

DÉPARTEMENT D'ÉCONOMIQUE

Faculté d'Administration

Université de Sherbrooke

**FORMULATION ET ESTIMATION DES MODÈLES ARCH ET GARCH AVEC
APPLICATION À L'ANALYSE DE LA VOLATILITÉ DES SÉRIES
ÉCONOMIQUES**

Par

GABIN OTHA-NDOUMBA 1974-

Bachelier ès sciences (économique)

de l'université de Sherbrooke

MÉMOIRE PRÉSENTÉ

pour obtenir

LA MAÎTRISE ÈS SCIENCES (ÉCONOMIQUE)

SHERBROOKE

JUIN 2004

I-2062

REMERCIEMENTS

Je tiens d'abord à exprimer toute ma reconnaissance à mon directeur M. Gérald ROY pour ses conseils et le soutien dont il a fait preuve tout au long de cette recherche, qui ont suscité en moi le désir d'apprendre.

Également, je tiens à remercier les membres de mon comité de lecture, M. Pene KALULUMIA et M. Mario FORTIN pour leur appui.

Je tiens enfin à exprimer ma profonde reconnaissance à mes parents, ma fiancée, mon fils, et à tous ceux que ma réussite importe.

TABLE DES MATIÈRES

REMERCIEMENTS	I
TABLE DES MATIÈRES.....	II
LISTE DES FIGURES	IV
LISTE DES TABLEAUX	V
INTRODUCTION GÉNÉRALE.....	1
CHAPITRE 1	4
FORMULATION	4
1.1 INTRODUCTION.....	4
1.2 LA FAMILLE DES MODÈLES ARCH	4
1.2.1 Le modèle ARCH (q).....	6
1.2.2 Le modèle GARCH (p,q).....	8
1.2.3 Propriétés des modèles (G)ARCH	10
1.2.3.1 Propriétés sur la loi de distribution normale	11
1.2.3.2 Loi des erreurs	16
ESTIMATION ET TEST.....	20
2.1 INTRODUCTION.....	20
2.2.1 La méthode d'estimation du maximum de vraisemblance (MV)	21
2.2.2 Estimation des modèles ARCH et GARCH	36
2.2.3 Les dérivées de la log-vraisemblance d'un modèle ARCH(1).....	39
2.2.4 Les dérivées de la log-vraisemblance d'un modèle GARCH(1,1)	41
2.3 TEST D'HOMOSCÉDASTICITÉ : TEST ARCH	44
2.4 TESTS D'ASYMÉTRIE ET D'APLATISSEMENT	45
CHAPITRE 3	51
APPLICATION EMPIRIQUE SUR LES RENDEMENTS BOURSIERS	51
3.1 INTRODUCTION.....	51
3.2 LA PREMIÈRE PARTIE : LES DONNÉES ET LEURS PROPRIÉTÉS STATISTIQUES	52

3.2.1	<i>Approche descriptive</i>	52
3.2.2	<i>Test de stationnarité : Test de KPSS</i>	59
3.2.3	<i>Test d'homoscédasticité</i>	61
3.2.4	<i>Test de normalité</i>	63
3.3	LA DEUXIÈME PARTIE : ESTIMATION ET SPÉCIFICATION	65
3.3.1	<i>Estimation de modèles (G)ARCH</i>	66
3.3.2	<i>Estimation de modèles (G)ARCH : Comparaison des résultats de plusieurs logiciels (ECTS, RATS et EViews)</i>	73
CONCLUSION		76
ANNEXE		78
PROGRAMMES ECTS		78
PROGRAMMES RATS.....		89
DOCUMENTS DE TRAVAIL ET ARTICLES TROUVÉS SUR INTERNET		95
BIBLIOGRAPHIE		96

LISTE DES FIGURES

3.1	Évolution temporelle des titres.....	54
3.2	Y_t vs Y_{t-1}.....	56
3.3	Fonctions de répartition empiriques et théoriques.....	58
3.4	Corrélogrammes des séries.....	60
3.5	Résidus au carré et var. conditionnelle du GARC(1,1).....	71

LISTE DES TABLEAUX

3.1	Statistiques descriptives des séries mensuelles.....	55
3.2	Test de stationnarité.....	62
3.3	Statistiques du test d'homoscédasticité.....	63
3.4	Corrélogrammes des résidus au carré.....	63
3.5	Résultats du test de normalité.....	65
3.6	Test d'autocorrélation résiduelle.....	69
3.7	Résultats de la régression OPG.....	70
3.8	Diagnostic de corrélation sérielle.....	73
3.9	Résultats comparatifs.....	75

Introduction générale

Le développement des modèles ARCH se place dans le contexte et la lignée des modèles traditionnels de séries chronologiques. Ces modèles pour séries temporelles ont été développés dans une perspective de description, de désaisonnalisation, de prévision ou de contrôle de systèmes. L'âge d'or de cette modélisation se situe dans les années soixante-dix, qui ont vu le développement des modèles autorégressifs à moyennes mobiles (ARMA) et leur généralisation.

Ces modèles sont fondés sur une approximation linéaire de la valeur présente d'une série en fonction linéaire de ses valeurs passées et de la valeur présente d'un bruit blanc (BB), s'interprétant comme l'innovation de la série. Plus précisément, on dit qu'un processus stationnaire X_t suit un ARMA(p,q) s'il vérifie la relation suivante :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

où les coefficients $\phi_i (i = 1, \dots, p)$ et $\theta_j (j = 1, \dots, q)$ sont des réels et $\varepsilon_t \approx BB(0, \sigma_\varepsilon^2)$. Cette formulation présente cependant quelques inconvénients. Elle est linéaire, ce qui restreint sa généralité. En outre, sa mise en œuvre impose peu de contraintes sur les paramètres autorégressifs, ce qui nuit à des interprétations plus structurelles des problèmes étudiés.

Parmi les domaines d'application où la formulation ARMA classique se révèle insuffisante figurent les problèmes financiers et monétaires. Les séries financières observables en pratique présentent d'abord certaines caractéristiques de dynamique non linéaire, dont la plus importante est le fait que la variabilité instantanée de la série (ou volatilité) dépend de façon importante du passé. Par

ailleurs, il existe des théories financières ou monétaires basées sur des principes d'équilibre et de comportements rationnels des agents intervenant sur le marché qui conduisent naturellement à introduire et à tester des contraintes structurelles sur les paramètres.

Les modèles ARCH (Autoregressive Conditional Heteroskedasticity) fournissent un cadre mieux adapté à ce genre de situation. Et depuis leur introduction, de nombreuses publications s'y rapportent. Cet engouement reflète l'importance accordée par la communauté scientifique à ce type de modèle, tant au point de vue statistique que financier.

Si l'on se place d'un point de vue statistique, les modèles ARCH constituent une classe spécifique de modèles non linéaires pour laquelle on peut mener une étude complète en abordant un certain nombre de problèmes classiques (test de marche aléatoire, loi des erreurs, calcul des moments, etc.).

D'un point de vue financier, la mise en œuvre de modèles où la dépendance temporelle de la volatilité joue un rôle fondamental conduit à repenser la théorie dans le même contexte. Parallèlement au développement des modèles ARCH, on assiste donc à la réécriture des modèles structurels concernant les déterminations de portefeuilles optimaux, les évaluations de prix des options, les liens entre des actifs de base et rendement du portefeuille de marché, etc.

Un survol de la littérature nous permet de constater que l'utilisation des modèles ARCH ne se limite plus à la théorie financière, car ils sont de plus en plus utilisés dans des buts décisionnels.

Ce mémoire est consacré à l'étude des modèles ARCH. Nous nous attarderons principalement sur les difficultés posées par l'estimation de ces modèles, qui selon des travaux antérieurs nécessitent l'utilisation de dérivées analytiques comme outil

d'estimation afin de pallier à la problématique du choix des valeurs initiales lors de l'estimation. Dans le chapitre I, nous présentons la formulation de ces modèles, ainsi que leurs propriétés. Puis brièvement, nous énumérons les différentes extensions de ces modèles.

Dans le chapitre II, nous abordons l'estimation des modèles ARCH par la méthode du maximum de vraisemblance. De nombreuses questions y sont particulièrement examinées : la difficulté posée par l'estimation de tels modèles, le calcul de dérivées analytiques nécessaires à l'estimation, la façon de mettre en œuvre des tests simples d'homoscédasticité, d'asymétrie et d'aplatissement.

Enfin, nous présentons au chapitre III un exemple simple d'application de la méthodologie ARCH à la modélisation des séries de rendements boursiers. Nous comparons ainsi les résultats obtenus par approches analytique et numérique. Les différentes étapes de la modélisation sont présentées en détail et illustrées à l'aide des logiciels ECTS, RATS et EVIEWS.

CHAPITRE 1

Formulation

1.1 Introduction

Dans ce chapitre, nous nous proposons de décrire les principales formulations des modèles avec hétéroscédasticité conditionnelle qui ont été proposés dans la littérature.

Nous commençons d'abord par présenter le modèle ARCH dans son ensemble et, brièvement, nous énumérerons les différentes extensions qui ont été faites sur ce modèle. Puis, nous étudierons les propriétés de ce modèle.

1.2 La famille des modèles ARCH

Les différentes études empiriques concernant les séries temporelles macro-économiques et financières ont montré que celles-ci étaient caractérisées par une distribution leptokurtique, une hétéroscédasticité corrélée à son passé. La spécification correcte de la variance conditionnelle est importante pour au moins deux raisons. D'une part, dans les études empiriques financières, l'aspect incertain des changements de prix est souvent mesuré par celle-ci; il est important pour les investisseurs de saisir le degré de risque puisque les agents ont une aversion face au risque et exigent, généralement, un rendement espéré supérieur en compensation d'un plus grand risque. D'autre part, les changements de la variance inter-temporelle ont d'importantes conséquences sur l'efficacité des estimateurs des

paramètres. Pour des variables tels que les taux de change, leur niveau apporte une information importante mais aussi leur volatilité, comme pour l'établissement du prix des options, et peut affecter le commerce international, voir F. Klaassen (1998).

Pourtant, les modèles économétriques conventionnels ne se concentrent pas sur les premiers moments (c'est le cas des modèles ARMA), et ignorent une quelconque dépendance de ceux plus élevés et supposent une variance constante.

Le processus ARCH, Autoregressive Conditional Heteroskedasticity, introduit par R.F. Engle (1982), permet à la variance conditionnelle de fluctuer dans le temps et d'être fonction du passé des erreurs tout en maintenant une variance inconditionnelle (globale) constante. Il apporte une meilleure description des mouvements de rendement des titres et d'autres séries macro-économiques et financières.

Ces dernières années, les modèles ARCH ont fait l'objet de beaucoup de discussions. Des publications telles que celle de T. Bollerslev, R.F. Engle et D.B. Nelson (1993) et A.K. Bera et M.L. Higgins (1993) présentent les différents avancements observés et suggèrent divers modèles et tests.

Selon la notation de T. Bollerslev, R.Y. Chou et K.F. Kroner (1992), le modèle ARCH se réfère à tous les processus temporels discrets $\{u_t\}$ de la forme $u_t = \varepsilon_t h_t$ où $\varepsilon_t \text{ iid}, E(\varepsilon_t) = 0, \text{var}(\varepsilon_t) = 1$, avec h_t une fonction de variance (scédastique), indépendante de ε_t .

T. Bollerslev, R.F. Engle et D.B. Nelson (1993) présentent quelques applications de ces modèles en finance où les rendements d'actifs tendent à être leptokurtiques ;

justement les modèles ARCH permettent de prendre en compte les grappes de volatilité, c'est-à-dire, les fortes (faibles) variations de prix qui sont suivies par d'autres fortes (faibles) variations de prix, mais dont le signe n'est pas prévisible, voir B. Mandelbrot (1963).

1.2.1 Le modèle ARCH (q)

Le modèle ARCH a été introduit dans le but de mieux analyser la dynamique des actifs financiers. L'article de base est celui de R.F. Engle (1982) qui fut le premier à proposer une paramétrisation possible de la variance. L'idée de base, ici, est que la variance du terme d'erreur à la période t dépend des valeurs réalisées du carré des erreurs dans les périodes précédentes.

Cette paramétrisation de la fonction de variance, h_t , s'exprime comme suit :

$$\sigma_t^2 \equiv E(u_t^2 | \Omega_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 = \alpha_0 + \alpha(L)u_{t-1}^2 \quad (1.9)$$

avec $\alpha_0 > 0$ et $\alpha_i \geq 0$, L représentant l'opérateur retard. Ce modèle est connu sous le nom de modèle ARCH(q). L'intérêt de cette approche réside dans l'introduction d'une dynamique endogène au niveau de la définition de la variance. Il fournit une spécification originale de l'hétéroscédasticité dont les causes sont inconnues.

L'ensemble de l'information disponible à la date $(t-1)$ est :

$$(\Omega_t = \{u_{t-1}, u_{t-2}, \dots, u_{t-q}\})$$

La variance conditionnelle à Ω_t , notée σ_t^2 , varie au cours du temps; elle est représentée par l'équation (1.9). Comme la variance conditionnelle de u_t est

fonction de u_{t-1} , ce qui implique que u_t et u_{t-1} sont dépendants, ils sont cependant non corrélés : $E(u_t u_{t-1}) = E(E(u_t u_{t-1} | \Omega_{t-1})) = 0$.

Il existe donc une quantité d'informations entre l'indépendance et la non-corrélation, qui peut être exploitée.

Par un raisonnement analogue, nous pouvons montrer que $E(u_t u_s) = 0$ pour tout $s < t$. Ceci implique qu'il y a seulement présence d'hétéroscédasticité dans le processus ARCH, mais pas de corrélation sérielle.

Si le modèle ARCH est stationnaire, alors la variance non conditionnelle des innovations, notée σ^2 , qui est l'espérance non conditionnelle de u_t^2 est indépendante du temps. Sous l'hypothèse de stationnarité, l'espérance non conditionnelle de l'équation (1.9) nous donne :

$$\sigma^2 = \alpha_0 + \sigma^2 \sum_{i=1}^q \alpha_i$$

Donc,

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i} \quad (1.10)$$

Elle est positive si $\alpha_i > 0$, pour $i = 0, 1, \dots, q$, et stationnaire si $\sum_{i=1}^q \alpha_i < 1$. Ce sont les contraintes de positivité.

La variance est estimée comme une moyenne mobile des carrés des innovations. Contrairement aux solutions traditionnelles qui spécifient le modèle uniquement en fonction des variables observables, les modèles ARCH prennent en compte directement la variation temporelle de la variance conditionnelle du processus du phénomène étudié dans l'estimation des paramètres d'intérêts.

En pratique, plusieurs séries financières présentent de la volatilité fluctuante dans le temps, mais avec une corrélation faible entre des valeurs de u_t^2 . Pour satisfaire ces deux caractéristiques, il devient nécessaire d'avoir un paramètre q élevé. Dans ce cas, il faut estimer plusieurs paramètres et la contrainte de positivité n'est pas toujours respectée.

1.2.2 Le modèle GARCH (p,q)

Pour pallier aux inconvénients ci-dessus, T. Bollerslev (1986) introduit le modèle GARCH (Generalized Autoregressive Conditional Heteroskedasticity), une généralisation du modèle ARCH dans laquelle sont ajoutés p retards de la variance conditionnelle h_t . Ce modèle permet une plus grande flexibilité dans la structure des retards. Il s'agit d'une extension du processus ARCH, et constitue un modèle ARMA de la variance conditionnelle alors que ARCH est un modèle AR de cette même variance. Toutefois, R.F. Engle (1995) observe que le modèle ARCH est une moyenne mobile puisque la variance conditionnelle est moyenne mobile des résidus au carré.

Le modèle GARCH (p,q) est tel que

$$h_t = \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^q \delta_j h_{t-j} = \alpha_0 + \alpha(L)u_{t-1}^2 + \delta(L)h_{t-1}, \quad (1.11)$$

ou encore :

$$\begin{aligned} [1 - \delta(L)]h_t &= \alpha_0 + \alpha(L)u_{t-1}^2 \\ \Leftrightarrow \Phi(L)y_t &= \gamma + \Theta(L)\varepsilon_t \end{aligned}$$

Tous les coefficients α_i et δ_j doivent être positifs et $\alpha_0 > 0$. Pour que le modèle soit stationnaire, il est recommandé d'avoir

$$\sum_{v=1}^p \delta_v + \sum_{s=1}^q \alpha_s < 1.$$

Prenons le cas d'un GARCH(1,1), où la variance conditionnelle s'écrit :

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1 \sigma_{t-1}^2. \quad (1.12)$$

Sous l'hypothèse de stationnarité, la variance non conditionnelle, σ^2 , peut être trouvée en prenant l'espérance non conditionnelle de (1.12). Nous obtenons :

$$\sigma^2 = \alpha_0 + \alpha_1 \sigma^2 + \delta_1 \sigma^2.$$

Donc,

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \delta_1}. \quad (1.13)$$

Pour que cette variance existe et qu'elle soit positive, il faut que $\alpha_1 + \delta_1 < 1$, et $\alpha_0 > 0$.

Puisque le modèle GARCH (p,q) correspond à un modèle ARCH de dimension infinie ARCH (∞), la plupart du temps un modèle GARCH (1,1), GARCH (1,2) ou GARCH (2,1) est employé. T. Bollerslev, R.Y. Chou et K.F. Kroner (1992) notent qu'un modèle GARCH (p,q) avec p et q petits peut approcher un modèle ARCH (q) où q est grand et modéliser les dynamiques de la variance sur de longues périodes.

Différentes extensions ont été proposées à ces modèles. Dans le modèle ARCH-M de Engle, Lilien et Robins (1987), la moyenne conditionnelle est fonction de la variance conditionnelle, voir T. Bollerslev, R.F. Engle et D.B. Nelson (1993). Le modèle E-GARCH introduit par D.B. Nelson (1991) modélise le log de la variance conditionnelle. Il se distingue des modèles (G)ARCH usuels par le fait qu'il rejette l'hypothèse de symétrie liée à la spécification quadratique de la variance conditionnelle. Face à la persistance de la volatilité, Engle et Bollerslev (1986) ont

proposé le modèle GARCH intégré, IGARCH, pour incorporer les résultats empiriques $\hat{\alpha} + \hat{\delta}_1$ est comprise entre 0.9 et 1, dans le cas d'un GARCH (1,1).

On retrouve aussi des modèles (G)ARCH multivariés. En effet, dans la plupart des cas, l'évaluation des actifs ou la décision d'allocation des portefeuilles doit être étudiée dans un contexte multivarié pour être significative. Ainsi u_t désignera maintenant un processus stochastique vectoriel. La notion de modèle ARCH est élargie à une représentation de la forme:

$$u_t = \varepsilon_t \Gamma_t^{1/2}$$

$$\varepsilon_t \text{ iid } E(\varepsilon_t) = 0 \quad \text{Var}(\varepsilon_t) = I$$

où la matrice $N \times N$ de la variance covariance Γ_t est définie positive et mesurable par rapport à l'ensemble de l'information de la date $t - 1$. La définition générale des processus multivariés définis par les deux équations ci-dessus autorise une grande variété de représentations possibles, mais peu d'entre elles sont utiles. Dans la représentation ARCH(q) multivariée proposée par Kraft et Engle (1983) Γ_t est une fonction linéaire des produits matriciels des carrés des erreurs passées. Ce modèle a été généralisé par la suite au modèle GARCH(p, q) par Bollerslev, Engle et Wooldridge (1988).

1.2.3 Propriétés des modèles (G)ARCH

Dans les sous-sections précédentes, nous avons brièvement évoqué certaines propriétés, notamment les conditions d'existence des moments.

Dans ce qui suit, nous allons étudier plus en détail les propriétés portant sur les moments de ces modèles.

1.2.3.1 Propriétés sur la loi de distribution normale

La fonction de densité d'une distribution normale centrée réduite est notée

$$\phi(x) = f(x) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right) \quad (1.14)$$

Une propriété fondamentale d'une variable aléatoire (v.a) X est son espérance mathématique qui est définie par

$$E(X) \equiv \int_{-\infty}^{+\infty} x\phi(x)dx. \quad (1.15)$$

Lorsque l'on parle d'espérance mathématique d'une v.a, on se réfère souvent à son premier moment. Les moments d'ordre supérieur, s'ils existent, sont aussi les espérances mathématiques de la v.a élevée à une certaine puissance. Ainsi, le second moment de la v.a X est l'espérance mathématique de X^2 , le 3^{ième} moment est l'espérance de X^3 , ainsi de suite. De manière générale, le $k^{\text{ième}}$ moment d'une variable aléatoire, X , continue est donnée par

$$m_k(X) \equiv \int_{-\infty}^{+\infty} x^k f(x)dx. \quad (1.16)$$

Afin de mieux saisir les propriétés ci-dessus, calculons les deux premiers moments de la distribution normale.

Les dérivées première et seconde de (1.14), par apport à x , sont respectivement :

$$\begin{aligned} \phi'(x) &= -(2\pi)^{-1/2} 2 \frac{1}{2} x \exp\left(-\frac{1}{2}x^2\right) \\ &= -x(2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right) = -x\phi(x), \end{aligned} \quad (1.17)$$

ou encore :

$$x\phi(x) = -\phi'(x), \quad (1.18)$$

et

$$\begin{aligned} \phi''(x) &= -\phi'(x) + x(2\pi)^{-1/2} 2 \frac{1}{2} x \exp\left(-\frac{1}{2}x^2\right) \\ &= -\phi(x) + x^2\phi(x). \end{aligned} \quad (1.19)$$

Ainsi, le premier moment de la distribution est égal à :

$$\begin{aligned} m_1(X) = \mu &= \int_{-\infty}^{+\infty} (2\pi)^{-1/2} x \exp\left(-\frac{1}{2}x^2\right) dx = \int_{-\infty}^{+\infty} x(2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right) dx = \int_{-\infty}^{+\infty} x \phi(x) dx. \\ &= - \int_{-\infty}^{+\infty} \phi'(x) dx = -[\phi(x)]_{-\infty}^{+\infty} = 0, \text{ à cause de la symétrie de la loi normale} \end{aligned}$$

par rapport à 0.

Et par intégration par parties, c'est-à-dire :

$$\int u dv = \int u dv - \int v du$$

$$\text{où } u = x, dv = x \exp\left(-\frac{x^2}{2}\right), \int dv = -\exp\left(-\frac{x^2}{2}\right), \text{ et } \int v du = \int \exp\left(-\frac{x^2}{2}\right) dx,$$

le deuxième moment (la variance) est égal à :

$$\begin{aligned} m_2(x) = \mu_2 &= \int_{-\infty}^{+\infty} x^2 f(x) dx = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx \\ &= (2\pi)^{-1/2} \left(-x \exp\left(-\frac{x^2}{2}\right) \right) \Big|_{-\infty}^{+\infty} + (2\pi)^{-1/2} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx, \end{aligned}$$

le membre de gauche, de l'expression ci-dessus, est égal à 0, nous obtenons alors

$$m_2(x) = \mu_2 = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1.$$

Ces deux résultats caractérisent les deux premiers moments d'une distribution normale centrée réduite. Dans cet exemple, nous avons constaté que le 1^{er} moment est égal à 0, alors que le 2^e moment est égal à 1. Ces résultats résultent du fait que,

par symétrie de la loi, les moments impairs sont nuls et les moments pairs sont non nuls. Vérifions cette propriété. Pour cela, nous utilisons l'expression (1.14).

Ainsi, pour les moment d'ordre pair, nous avons n , un entier naturel, $2n$ est pair.

Soit μ_{2n} un moment d'ordre pair. Par définition :

$$\begin{aligned} m_{2n} &= \mu_{2n} = E(x)^{2n} \\ &= \int_{-\infty}^{+\infty} x^{2n} \phi(x) dx = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} x^{2n} \exp\left(-\frac{x^2}{2}\right) dx. \end{aligned}$$

Soit $M(t)$, la fonction génératrice des moments. Elle est définie par

$$M(t) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} \exp(tx) \exp\left(-\frac{x^2}{2}\right) dx,$$

alors $M(t) = \exp\left(\frac{1}{2}t^2\right).$

$\forall t \in$ de l'ensemble des réels, $\exp(tx) \exp\left(-\frac{x^2}{2}\right) = \sum_{n=0}^{\infty} \left(\frac{t^n}{n!} x^n \exp\left(-\frac{x^2}{2}\right) \right)$, série qui converge uniformément en x .

Nous avons alors :

$$M(t) = \sum_{n=0}^{\infty} \left(\frac{t^n}{n!} \mu^n \right).$$

Ainsi, le moment μ_n est le coefficient de $\frac{t^n}{n!}$ dans $\exp\left(\frac{1}{2}t^2\right)$, et

$$\mu_{2n} = \frac{1}{2^n} \frac{(2n)!}{n!}. \quad (1.20)$$

À partir de l'expression (1.20), nous pouvons, par exemple, tirer le 4^e moment de la distribution,

$$n = 2.$$

Il vient alors :

$$\mu_{2 \times 2} = \frac{1}{2^2} \frac{(2 \times 2)!}{2!} = \frac{1}{2^2} \frac{4 \times 3 \times 2!}{2!} = 3.$$

Dans le cas où, le 2^e moment est égal à σ^2 , l'expression (1.20) s'écrit :

$$\mu_{2n} = \frac{\sigma^2}{2^n} \frac{(2n)!}{n!}. \quad (1.21)$$

Il s'agit de la généralisation de l'expression (1.20), avec fonction de densité $\phi(x)$ définie comme suit :

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} x^2\right).$$

Pour les moments d'ordre impair, nous posons :

$$\mu_{2n+1} = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} x^{(2n+1)} \exp\left(-\frac{x^2}{2}\right) dx.$$

Montrons que cette intégrale est égale à 0. Pour ce faire, posons :

$$F_n(x) = -x^{(2n+1)} \exp\left(-\frac{x^2}{2}\right).$$

Étudions la parité de $F_n(x)$, c'est-à-dire :

$$\begin{aligned} F_n(-x) &= -x^{(2n+1)} \exp\left(-\left(-\frac{x}{2}\right)^2\right) = (-1)^{(2n+1)} x^{(2n+1)} \exp\left(-\frac{x^2}{2}\right). \\ &= -x^{(2n+1)} \exp\left(-\frac{x^2}{2}\right) = -F(x). \end{aligned}$$

La fonction F_n est alors dite impaire.

Donc,

$$\int_{-\infty}^{+\infty} F_n(x) dx = 0$$

Nous pouvons donc déduire que les moments d'ordre impair, d'une distribution normale, sont nuls. Par exemple le 3^e moment, où $n = 1$, est déterminé comme suit :

$$\mu_{2 \times 1 + 1} = \mu_3 = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} x^{(2 \times 1 + 1)} \exp\left(-\frac{x^2}{2}\right) dx = 0.$$

Nous avons montré que les moments d'ordre pair d'une distribution normale sont tirés de l'expression (1.21).

Cette expression est équivalente à $\prod_{j=1}^n (2j-1)$.

Montrons que :

$$\mu_{2n} = \frac{1}{2^n} \frac{(2n)!}{n!} = \prod_{j=1}^n (2j-1). \quad (1.22)$$

Utilisons un raisonnement par induction sur n .

Pour $n=1$, vérifions l'égalité (1.22).

$$\mu_2 = \frac{2!}{2 \times 1!} = 1,$$

et

$$\prod_{j=1}^1 (2j-1) = 2 \times 1 - 1 = 1.$$

L'égalité est vérifiée au rang $n=1$,

Hypothèse d'induction :

Supposons que $\frac{(2n)!}{2^n n!} = \prod_{j=1}^n (2j-1)$.

Vérifions que cette égalité est vraie au rang $n+1$.

Nous avons alors :

$$\begin{aligned} \frac{(2(n+1))!}{2^{(n+1)}(n+1)!} &= \frac{(2n+2)(2n+1)(2n)!}{2^n 2(n+1)n!} \\ &= \frac{(2n)!}{2^n n!} \times \frac{(2n+2)(2n+1)}{2(n+1)} = \frac{(2n)!}{2^n n!} \times (2n+1) \\ &= \prod_{j=1}^n (2j-1) \times (2n+1). \\ &= \prod_{j=1}^{n+1} (2j-1). \end{aligned}$$

Donc $\forall n$, nous avons :

$$\frac{(2n)!}{2^n n!} = \prod_{j=1}^n (2j-1). \quad \text{C.Q.F.D.}$$

Nous pouvons donc conclure que le $(2n)^{ième}$ moment d'une distribution normale centrée réduite est égale à

$$\prod_{j=1}^n (2j-1). \quad (1.23)$$

Ces résultats intermédiaires vont maintenant nous servir à caractériser la distribution des erreurs dans les modèles G(ARCH).

1.2.3.2 Loi des erreurs

Dans le cas d'un processus ARMA, la distribution non conditionnelle de u_t est habituellement normale, en autant que ε_t soit bruit blanc. Mais lorsque les erreurs sont de type (G)ARCH, ceci n'est pas valide. Comme nous le verrons, la distribution stationnaire des modèles (G)ARCH n'est pas normale, et les moments d'ordre 1 et 2 peuvent ne pas exister.

Afin de déterminer les deuxième et quatrième moments de ces processus, nous allons d'abord porter notre attention sur le modèle ARCH(1).

Pour un processus ARCH(1) dont les innovations sont normalement distribuées et suivent un bruit blanc, la distribution de u_t conditionnellement à Ω_t est normale.

Comme la variance stationnaire de cette distribution est égale à σ^2 , les deuxième et quatrième moments valent respectivement σ^2 et $3\sigma^4$, $\sigma^2 = \alpha_0 + \alpha_1 u_{t-1}^2$. Donc.

$$E(u_t^2 | \Omega_{t-1}) = \sigma^2 = \alpha_0 + \alpha_1 u_{t-1}^2.$$

et

$$E(u_t^4 | \Omega_{t-1}) = 3(\alpha_0 + \alpha_1 u_{t-1}^2)^2 = 3\alpha_0^2 + 6\alpha_0\alpha_1 u_{t-1}^2 + 3\alpha_1^2 u_{t-1}^4.$$

Si nous supposons que les deuxième et quatrième moments non conditionnels existent et nous notons le quatrième moment m_4 , en prenant l'espérance non conditionnelle des deux relations ci-dessus, nous obtenons

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1} \quad (1.24)$$

et

$$m_4 = 3\alpha_0^2 + \frac{6\alpha_0^2\alpha_1}{1 - \alpha_1} + 3\alpha_1^2 m_4,$$

où nous avons utilisé le résultat de la première équation, celle du 2^{ème} moment, pour déduire la 2^{ème} équation. En résolvant cette équation par rapport à m_4 , nous obtenons

$$m_4 = \frac{3\alpha_0^2(1 - \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}. \quad (1.25)$$

Ces résultats ne sont valides que si $\alpha_1 < 1$ et $3\alpha_1^2 < 1$. En effet, si ces conditions ne sont pas satisfaites, les 2^{ème} et 4^{ème} moments n'existent pas.

Engle (1982) a introduit son modèle en supposant un processus u_t avec une distribution conditionnellement normale :

$$u_t | \Omega_{t-1} \approx N(0; h_t). \quad (1.26)$$

Il est clair que ce processus conditionnellement gaussien n'est pas marginalement gaussien; en effet si c'était le cas, la variance conditionnelle devrait être indépendante des valeurs passées, condition non réalisée.

Des résultats plus précis peuvent être obtenus concernant l'aspect non gaussien de la distribution marginale de u_t .

En effet, d'après les résultats (1.25) et (1.26), nous pouvons déduire que le coefficient d'aplatissement, kurtosis, associé à la distribution marginale vaut

$$Ku = \frac{E(u_t^4)}{E(u_t^2)^2} = 3 \frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3 \frac{1 - \alpha_1^2}{(1 - 3\alpha_1^2)} > 3. \quad (1.27)$$

Ce coefficient d'aplatissement est toujours supérieur à 3, valeur correspondant à la loi normale. Pour cette raison, la loi des erreurs est dite *leptokurtique*.

Dans le cas d'un GARCH (1,1), la fonction scédastique est

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1 \sigma_{t-1}^2. \quad (1.28)$$

La variance stationnaire vérifie l'équation

$$\begin{aligned} \sigma^2 &= \alpha_0 + (\alpha_1 + \delta_1) \sigma^2 \\ &= \frac{\alpha_0}{1 - \alpha_1 - \delta_1}. \end{aligned}$$

Soit m_4 l'espérance non conditionnelle de σ_t^4 . Nous avons alors :

$$\begin{aligned} m_4 &= E(\alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1 \sigma_{t-1}^2)^2 \\ &= E[\alpha_0^2 + \alpha_1^2 u_{t-1}^4 + \delta_1^2 \sigma_{t-1}^4 + 2\alpha_0(\alpha_1 u_{t-1}^2 + \delta_1 \sigma_{t-1}^2) + 2\alpha_1 \delta_1 u_{t-1}^2 \sigma_{t-1}^2]. \end{aligned}$$

Or $E(\sigma_{t-1}^4) = E(\sigma^4) = E(\sigma_1^4) = m_4$.

Comme $u_t = \sigma_t \varepsilon_t$ avec $\varepsilon_t \approx N(0,1)$, on a que $u_{t-1}^4 = \sigma_{t-1}^4 \varepsilon_{t-1}^4$ et $u_{t-1}^2 = \sigma_{t-1}^2 \varepsilon_{t-1}^2$.

Donc,

$$E(u_{t-1}^4) = 3m_4 \text{ et } E(u_{t-1}^2 \sigma_{t-1}^2) = m_4.$$

Les espérances non conditionnelles de u_{t-1}^2 et σ_{t-1}^2 sont σ^2 . Alors

$$m_4 = 3\alpha_1^2 m_4 + \delta_1^2 m_4 + 2\alpha_1 \delta_1 + \alpha_0^2 + 2\alpha_0(\alpha_1 + \delta_1) \sigma^2.$$

Or $\frac{\alpha_0}{1 - \alpha_1 - \delta_1}$, ce qui nous donne

$$m_4(1 - 3\alpha_1^2 - \delta_1^2 - 2\alpha_1 \delta_1) = \alpha_0^2 + 2\alpha_0^2 \frac{(\alpha_1 + \delta_1)}{1 - \alpha_1 - \delta_1}$$

$$\Rightarrow m_4(1 - (\alpha_1 + \delta_1)^2 - 2\alpha_1^2) = \alpha_0^2 \frac{1 + \alpha_1 + \delta_1}{1 - \alpha_1 - \delta_1},$$

$$\text{d'où } m_4 = \alpha_0^2 \frac{(1 + \alpha_1 + \delta_1)(1 - \alpha_1 - \delta_1)}{1 - (\alpha_1 + \delta_1)^2 - 2\alpha_1^2}. \quad (1.29)$$

On a vu que le 4^e moment non conditionnel de u_t est $3m_4$. Nous avons alors

$$E(u_{t-1}^4) = 3 \frac{\alpha_0^2(1 + \alpha_1 + \delta_1)(1 - \alpha_1 - \delta_1)}{1 - (\alpha_1 + \delta_1)^2 - 2\alpha_1^2}. \quad (1.30)$$

Pour que ces moments existent, il faut que le dénominateur de l'équation de m_4 soit positif. Si $\delta_1 = 0$, cette condition devient

$$1 - \alpha_1^2 - 2\alpha_1^2 = 1 - 3\alpha_1^2 > 0,$$

ce qui équivaut à la condition $3\alpha_1^2 < 1$, qui est la condition d'existence du 4^e moment non conditionnel d'un ARCH(1).

Une fois de plus, on déduit que la loi marginale de u_t a des queues plus lourdes qu'une loi normale. Nous voyons qu'il y a accroissement de l'aplatissement lorsqu'on passe des lois conditionnelles aux lois non conditionnelles.

CHAPITRE 2

Estimation et test

2.1 Introduction

Les modèles introduits dans le chapitre précédent reposent sur des formulations des moyennes et des variances conditionnelles. En pratique, celles-ci sont souvent exprimées de façon à ce qu'elles apparaissent comme des fonctions de paramètres inconnus et des valeurs passées du processus. La connaissance de ces moments ne suffit cependant pas, sans l'hypothèse supplémentaire de la forme de la distribution, à caractériser la loi conditionnelle du processus. Le maximum de vraisemblance (MV) sera la méthode d'estimation utilisée dans ce mémoire.

Dans cette partie, nous nous attacherons à présenter les notions statistiques nécessaires à l'estimation et aux tests. Nous présenterons, en effet, les éléments requis à l'estimation de ces modèles.

De façon à ne pas compliquer la présentation et parce que les résultats ne sont pas spécifiques aux modèles ARCH-GARCH, nous n'insisterons pas sur les démonstrations de certaines propriétés asymptotiques, ni sur les conditions de régularité sous lesquelles elles sont valides.

2.2.1 La méthode d'estimation du maximum de vraisemblance (MV)

L'idée fondamentale de l'estimation par maximum de vraisemblance est, comme le nom l'indique, de trouver un ensemble d'estimations de paramètres tel que la vraisemblance d'avoir obtenu l'échantillon que nous utilisons soit maximisée. Nous signifions par là que la fonction de densité jointe (FDC) pour le modèle que l'on estime est évaluée aux valeurs observées de la (des) variable(s) et traitée(s) comme une fonction de paramètres du modèle. Si on peut simuler la variable dépendante, cela signifie que la FDC doit être connue, à la fois pour chaque observation comme variable scalaire aléatoire, et pour l'ensemble de l'échantillon comme un vecteur de variables aléatoires.

Comme d'habitude, nous décrivons la variable dépendante par un vecteur y . Pour un vecteur k donné de paramètres θ , la fonction de densité conjointe de y s'écrit $f(y, \theta)$. Cette FDC constitue la spécification du modèle. Puisqu'une fonction de densité de probabilité fournit un contenu ambigu pour la simulation, il suffit, dans ce cas, de spécifier le vecteur θ afin de donner la caractérisation complète du processus générateur des données (PGD). Il y a aussi une relation de correspondance biunivoque entre les PGDs du modèle et les vecteurs de paramètres admissibles.

L'estimation par maximum de vraisemblance est basée sur la spécification du modèle à travers la fonction de densité conjointe $f(y, \theta)$. Quand θ est fixé, la fonction $f(\cdot, \theta)$ de y est interprétée comme la fonction de densité de y . Mais si $f(y, \theta)$ est plutôt évaluée par le vecteur y trouvé dans un ensemble de données, la fonction $f(y, \cdot)$ des paramètres du modèle ne peut plus être, alors interprétée

comme une FDC. Cela renvoie plutôt à une fonction de vraisemblance du modèle pour l'ensemble des données spécifiées. L'estimation par maximum de vraisemblance (MV) revient alors à maximiser la fonction de vraisemblance. Le vecteur de paramètres $\hat{\theta}$ auquel la vraisemblance atteint sa valeur maximale est appelé estimateur du maximum de vraisemblance ou EMV, des paramètres.

Dans bien des cas, les observations successives d'un échantillon sont autre que le produit des densités des observations individuelles. Posons $f(y_i, \theta)$ la FDC d'une observation, y_i .

La densité jointe de tout l'échantillon y s'écrit alors :

$$f(y, \theta) = \prod_{i=1}^n f(y_i, \theta) \quad (2.1)$$

supposées être statistiquement indépendantes. Dans ce cas, la densité jointe de l'échantillon n'est nul

Dans le cas d'échantillons de grande taille, (2.1) peut devenir extrêmement importante ou extrêmement petite, et prendre des valeurs qui sont bien au-delà des capacités des ordinateurs. Pour cette raison, parmi d'autres, il est d'usage de maximiser le logarithme de la fonction de vraisemblance plutôt que la fonction de vraisemblance elle-même. Bien évidemment, nous obtiendrons la même réponse en procédant ainsi, car la **fonction logvraisemblance** est une transformation monotone de la fonction de vraisemblance. Nous écrivons alors :

$$l(y, \theta) \equiv \log f(y, \theta) = \sum_{i=1}^n l_i(y_i, \theta), \quad (2.2)$$

où $l_i(y_i, \theta)$, la **contribution** à la fonction logvraisemblance faite par chaque observation i , est égale au $\log f_i(y_i, \theta)$.

La manière la plus simple de saisir l'idée fondamentale de l'estimation par MV est de considérer un exemple simple. Supposons que chaque observation y_i soit générée par la fonction de densité

$$f(y_i, \theta) = \theta e^{-\theta y_i}, \quad y_i > 0, \quad \theta > 0 \quad (2.3)$$

et soit indépendante de toutes les autres y_i . Il s'agit de la fonction de densité de la **distribution exponentielle**. Il y a un seul paramètre inconnu θ que nous désirons estimer, et nous disposons de n observations avec lesquelles nous allons travailler. Comme les y_i sont indépendantes, leur densité jointe est simplement le produit de leurs densités marginales. La fonction de vraisemblance s'écrit alors

$$f(y, \theta) = \prod_{i=1}^n \theta e^{-\theta y_i}. \quad (2.4)$$

La fonction de logvraisemblance correspondant à (2.4) est

$$l(y, \theta) = \sum_{i=1}^n (\log(\theta) - \theta y_i) = n \log(\theta) - \theta \sum_{i=1}^n y_i. \quad (2.5)$$

La maximisation de la fonction de logvraisemblance par rapport au seul paramètre inconnu θ , est une procédure directe. Différencier l'expression plus à droite de (2.5) par rapport à θ et poser la dérivée à zéro donne la condition de premier ordre (CPO)

$$\frac{n}{\theta} - \sum_{i=1}^n y_i = 0, \quad (2.6)$$

et nous trouvons que l'estimateur MV $\hat{\theta}$ est

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n y_i} \quad (2.7)$$

Dans ce cas, il n'est pas nécessaire de se soucier des multiples solutions de (2.6). La dérivée seconde de (2.7) est toujours négative, ce qui nous permet de conclure que $\hat{\theta}$ défini par (2.7) est l'unique estimateur MV. Notons que cela ne sera pas

toujours le cas; pour certains problèmes les CPO peuvent mener à des solutions simples.

Après cet exemple, nous poursuivons notre présentation sur la méthode d'estimation par maximum de vraisemblance.

Lorsque la fonction de vraisemblance ne peut pas s'écrire sous la forme de (2.1), il est toujours possible, en théorie, de factoriser $f(y, \theta)$ en une série de contributions, chacune provenant d'une seule observation.

Supposons que les observations individuelles $y_t, t = 1, \dots, n$, soient ordonnées d'une certaine manière, comme dans le cas des séries temporelles où les observations sont dépendantes. Or cette factorisation peut être accomplie comme suit. Nous commençons par la densité marginale de la première observation y_1 , que l'on peut appeler $f_1(y_1)$, en supprimant la dépendance par rapport à θ pour le moment, et ce dans le but d'alléger l'écriture. Puis la densité conjointe des deux premières observations peut être écrite comme le produit de $f_1(y_1)$ par la densité de y_2 conditionnellement à y_1 , et nous la notons $f_{21}(y_2|y_1)$. Si maintenant, nous prenons les trois premières observations ensemble, leur densité jointe est le produit de la densité conditionnelle des deux premières prises simultanément, par la densité de la troisième conditionnellement aux deux premières, et ainsi de suite. Le résultat pour l'échantillon entier des observations est

$$\begin{aligned} f(y) &= f_1(y_1)f_2(y_2|y_1)f_3(y_3|y_2, y_1) \dots f_n(y_n|y_{n-1}, \dots, y_1) \\ &= \prod_{t=1}^n f_t(y_t|y_{t-1}, \dots, y_1) \end{aligned} \quad (2.8)$$

Notons que ce résultat (2.8) est parfaitement général et peut être appliqué à n'importe quelle densité ou fonction de vraisemblance. L'ordre des observations est

habituellement l'ordre naturel, comme pour les séries temporelles, mais même si aucun ordre naturel n'existe, l'équation (2.8) demeure vraie pour un classement arbitraire.

Comme nous l'avons précisé précédemment, dans la pratique on utilise la fonction de logvraisemblance $l(y, \theta)$ plutôt que la fonction de vraisemblance $f(y, \theta)$. La décomposition de $l(y, \theta)$ en contributions provenant d'observations individuelles résulte de l'équation (2.8). Elle peut être écrite comme suit, en supprimant la dépendance par rapport à θ pour alléger les notations :

$$l(y) = \sum l_i(y_i | y_{i-1}, \dots, y_1), \quad (2.9)$$

où $l_i(y_i | y_{i-1}, \dots, y_1) \equiv \log f_i(y_i | y_{i-1}, \dots, y_1)$, comme précédemment.

Nous pouvons dès à présent donner la définition de l'estimation par maximum de vraisemblance. Nous disons que $\hat{\theta}$ est une estimation par MV, pour les données y si

$$l(y, \hat{\theta}) \geq l(y, \theta) \quad \forall \theta \in \Theta, \quad (2.10)$$

Θ étant un espace paramétrique dans lequel θ prend ses valeurs.

Si l'inégalité est stricte, alors $\hat{\theta}$ est l'unique EMV, comme ce fut le cas précédemment dans notre exemple.

Il est souvent plus commode d'utiliser une autre définition de l'estimation par MV, qui n'est pas équivalente en général. Si la fonction de vraisemblance atteint un maximum intérieur à l'espace paramétrique, alors elle, ou de façon équivalente la fonction de logvraisemblance, doit satisfaire les CPO pour un maximum. Ainsi une estimation par MV peut se *définir* comme une solution aux équations de vraisemblance, qui correspondent précisément aux CPO suivantes:

$$g(y, \hat{\theta}) \equiv 0, \quad (2.11)$$

où $g(y, \hat{\theta})$ est le **vecteur gradient**, ou le **vecteur score**, dont les éléments sont :

$$g_i(y, \theta) \equiv \frac{\partial l(y, \theta)}{\partial \theta_i} = \sum_{i=1}^n \frac{\partial l_i(y_i, \theta)}{\partial \theta_i} \quad (2.12)$$

Comme nous l'avons signalé antérieurement, il peut arriver que plus d'une valeur de θ satisfasse les équations de vraisemblance (2.11), la définition nécessite par ailleurs que l'estimation $\hat{\theta}$ soit associée à un *maximum* local de $l(y, \theta)$ et que

$$p \lim_{n \rightarrow \infty} (n^{-1} l(y, \hat{\theta})) \geq p \lim_{n \rightarrow \infty} (n^{-1} l(y, \theta^*)),$$

où θ^* est n'importe quelle autre solution des équations de vraisemblance. Cette seconde définition de l'estimation par MV est souvent associée à Cramér, (Cramér, 1946). Dans la pratique, la nécessité que $p \lim(n^{-1} l(y, \hat{\theta})) \geq p \lim(n^{-1} l(y, \theta^*))$ est impossible à vérifier en général. Le problème vient du fait que l'on ne connaît pas le PGD et que par conséquent, le calcul analytique des limites de probabilité est impossible. Si pour un échantillon donné il existe deux racines ou plus aux équations de vraisemblance, celle qui associée à la valeur la plus haute de $l(y, \theta)$ pour cet échantillon peut ne pas converger vers celle qui est associée à la valeur la plus haute asymptotiquement. Dans la pratique, il existe plus d'une solution pour les équations de vraisemblance, l'on sélectionne celle qui est associé à la valeur la plus haute de la fonction de logvraisemblance.

Nous insistons sur le fait que ces deux définitions de l'estimation par MV ne sont pas équivalentes. En conséquence, il est parfois nécessaire de parler des estimations par **MV du type 1** quand nous faisons référence à celles obtenues par la maximisation de $l(y, \theta)$ sur Θ , et des estimations par **MV de type 2** quand nous faisons référence à celles obtenues comme solutions des équations de vraisemblance.

Par le terme **estimateur du maximum de vraisemblance** nous désignerons la variable aléatoire qui associe à chaque occurrence aléatoire possible y l'estimation par MV correspondante. Avant de poursuivre, rappelons la distinction entre une **estimation** et un **estimateur**. Nous pouvons rappeler qu'un estimateur, une variable aléatoire (v.a.), est représenté comme une fonction des ensembles possibles d'observations, alors qu'une estimation est une valeur que peut prendre cette fonction pour un ensemble d'observations bien spécifié.

Tout comme il existe deux définitions possibles des estimations MV, il existe également deux définitions possibles d'un estimateur MV. Les définitions suivantes montrent clairement que l'estimateur est une v.a. qui dépend des valeurs observées de l'échantillon y . L'estimateur de type 1, correspondant à la définition standard (2.10) de l'estimation MV, est $\hat{\theta}(y)$ défini par

$$l(y, \hat{\theta}(y)) > l(y, \theta) \quad \forall \theta \in \Theta \text{ tel que } \theta \neq \hat{\theta}(y). \quad (2.13)$$

L'estimateur de type 2, correspondant à la définition (2.11) de Cramér, est $\hat{\theta}(y)$ défini par :

$$g(y, \hat{\theta}(y)) = 0, \quad (2.14)$$

où $\hat{\theta}(y)$ donne le maximum local de l , et

$$p \lim_{n \rightarrow \infty} (n^{-1} l(y, \hat{\theta}(y))) \geq p \lim_{n \rightarrow \infty} (n^{-1} l(y, \theta^*(y))) \quad (2.15)$$

pour n'importe quelle autre solution $\theta^*(y)$ des équations de vraisemblance.

Nous concluons cette sous-section par une variété de définitions importantes pour la méthode d'estimation par MV. En utilisant la décomposition (2.9) de la fonction de logvraisemblance $l(y, \theta)$, nous pouvons définir une matrice $\mathbf{G}(y, \theta)$ de dimension $n \times k$ dont l'élément type est

$$G_{it}(y, \theta) \equiv \frac{\partial l_t(y, \theta)}{\partial \theta_i} \quad (2.16)$$

Étant donné que $g_i(y, \theta) \equiv \sum_{t=1}^n G_{it}(y, \theta)$. Nous appellerons $\mathbf{G}(y, \theta)$ la **matrice des contributions au gradient**, ou matrice **CG** tout simplement. Cette matrice est intimement reliée au vecteur gradient \mathbf{g} , qui est égale à $\mathbf{G}^T \mathbf{l}$, où \mathbf{l} désigne un vecteur de taille n constitué de 1 seulement et l'on note \mathbf{G}^T la transposée de \mathbf{G} . La $t^{i\text{ème}}$ ligne de \mathbf{G} , qui mesure la contribution au gradient de la $t^{i\text{ème}}$ observation, sera notée G_t .

La **matrice Hessienne** associée à la fonction de logvraisemblance $l(y, \theta)$ est la matrice $\mathbf{H}(y, \theta)$ de dimension $k \times k$ dont l'élément type est

$$\mathbf{H}_{ij}(y, \theta) \equiv \frac{\partial^2 l(y, \theta)}{\partial \theta_i \partial \theta_j}. \quad (2.17)$$

Nous définissons la **matrice Hessienne asymptotique**, si elle existe, comme

$$\mathbf{K}(\theta) \equiv p \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{H}(y, \theta)$$

Cette quantité, qui est une matrice symétrique, et en général semi-définie négative, apparaît souvent dans la théorie asymptotique de l'estimation par MV.

Nous définissons l'**information contenue dans l'observation t** par $\mathbf{I}_t(\theta)$, la matrice de dimension $k \times k$ dont l'élément type est

$$(\mathbf{I}(\theta))_{ij} \equiv E_\theta(G_{it}(\theta)G_{jt}(\theta)). \quad (2.18)$$

La notation " E_θ ", ici, signifie que l'espérance a été prise sous le PGD caractérisée par θ . Le fait que $\mathbf{I}_t(\theta)$ soit une matrice symétrique, en général semi-définie positive, et qu'elle soit définie positive à condition qu'il existe une relation linéaire entre les composantes du vecteur aléatoire \mathbf{G}_t est une conséquence immédiate de

cette définition. Nous pouvons alors définir la **matrice d'information** $\mathbf{I}(\theta)$, comme

$$\mathbf{I}(\theta) \equiv \sum_{t=1}^n \mathbf{I}_t(\theta) = \sum_{t=1}^n E_{\theta}(G_t(y, \theta)G_t(y, \theta)),$$

la matrice $\mathbf{I}(\theta)$ est symétrique et semi-définie positive.

La **matrice d'information asymptotique**, si elle existe, est définie par

$$\mathbf{J}(\theta) \equiv p \lim_{\theta, n \rightarrow \infty} \frac{1}{n} \mathbf{I}(\theta). \quad (2.19)$$

La matrice $\mathbf{I}_t(\theta)$ mesure la quantité *espérée* d'information contenue dans la $t^{ième}$ observation. La matrice d'information \mathbf{J} est, comme \mathbf{I}_t , symétrique et en général semi-définie positive.

Il existe une relation très importante entre la Hessienne asymptotique ($\mathbf{H}(\theta)$) et la matrice d'information asymptotique ($\mathbf{J}(\theta)$). Cette relation nous donne la **matrice de covariance asymptotique**, aussi appelée **égalité de la matrice d'information**, et nous la notons

$$\mathbf{K}(\theta) = -\mathbf{J}(\theta). \quad (2.20)$$

Littéralement, la matrice Hessienne asymptotique est l'opposé de la matrice d'information asymptotique. Formellement, nous pouvons donc dire que la matrice de covariance asymptotique de l'estimateur MV est donnée par

$$\hat{Var}(\hat{\theta}) = -\mathbf{K}^{-1}(\hat{\theta}). \quad (2.21)$$

ou par

$$\hat{Var}(\hat{\theta}) = \mathbf{J}^{-1}(\hat{\theta}). \quad (2.22)$$

L'obtention, à l'aide de (2.21) ou de (2.22), de $\hat{Var}(\hat{\theta})$ est parfois chose difficile dans certains modèles. C'est pourquoi, il existe une troisième méthode qui nous

permet d'obtenir l'estimateur de la matrice de covariance asymptotique. Il s'agit de l'estimateur **OPG (Produit-extérieur-du-gradient)**. Il est basé sur la définition

$$I(\theta) \equiv E(G^T(\theta_0)G(\theta_0))$$

ou θ_0 est un vecteur paramétrique associé au PGD.

Nous pouvons donc, logiquement, estimer $n^{-1}I(\theta_0)$ par $n^{-1}G^T(\hat{\theta})G(\hat{\theta})$. L'estimateur OPG est alors

$$\hat{Var}_{OPG}(\hat{\theta}) = (G^T(\hat{\theta})G(\hat{\theta}))^{-1} \quad (2.23)$$

C'est un estimateur que l'on calcule facilement. Et contrairement à la Hessienne empirique (2.21), il dépend seulement des dérivées premières. Et par rapport à (2.22), il ne nécessite pas de calculs théoriques. Toutefois, en échantillons finis, il est moins fiable que les deux autres.

Explicitons un peu plus la **régression OPG**, de laquelle découle l'estimateur OPG.

La régression OPG peut être utilisée pour n'importe quel modèle estimé par MV. Elle a été utilisée premièrement comme un moyen de calcul des statistiques de test par Godfrey et Wickens (1981). Il s'agit d'une régression artificielle qui permet de vérifier les CPO pour la fonction de logvraisemblance, l'estimation de la matrice de covariance, et ainsi de suite.

Supposons un modèle paramétré défini par la logvraisemblance (2.2). Soit $G(\theta)$ la matrice CG associée à la fonction de logvraisemblance (2.2), avec comme élément type

$$G_{it}(\theta) \equiv \frac{\partial l_t}{\partial \theta_i}; \quad t = 1, \dots, n, \quad i = 1, \dots, k,$$

où k est le nombre d'éléments dans le vecteur paramétrique θ . La régression OPG associée au modèle (2.2) peut alors être écrite comme

$$\iota = G(\theta)c + \text{résidus} \quad (2.24)$$

Ici, ι désigne encore un vecteur de taille n dans lequel chaque élément est unitaire et c est vecteur de dimension k des paramètres artificiels. Le produit de la matrice des régresseurs avec le régressé (ι) est le gradient $g(\theta) \equiv G^T(\theta)\iota$. La matrice des sommes des carrés et les produits croisés des régresseurs, $G^T(\theta)G(\theta)$, lorsqu'elle est divisée par n , estime de manière convergente la matrice d'information $J(\theta)$. Ces deux caractéristiques sont essentiellement tout ce qu'il faut pour que (2.24) soit une régression valide. Comme dans le cas d'une régression Gauss-Newton, les régresseurs de la régression OPG dépendent du vecteur θ . Par conséquent, avant d'effectuer la régression artificielle, ces régresseurs doivent être évalués pour un vecteur paramétrique spécifique.

Un choix possible pour ce vecteur paramétrique est $\hat{\theta}$ l'estimateur MV du modèle (2.2). Dans ce cas, la matrice du régresseur est $\hat{G} \equiv G(\hat{\theta})$ et les estimations paramétriques artificielles, qui sont désignées par \hat{c} , sont identiquement égales à zéro :

$$\hat{c} = (\hat{G}^T \hat{G})^{-1} \hat{G}^T \iota = (\hat{G}^T \hat{G})^{-1} \hat{g} = 0.$$

Comme ici \hat{g} est le gradient évalué en $\hat{\theta}$ de la fonction logvraisemblance, la dernière égalité ci-dessus est une conséquence des CPO pour le maximum de vraisemblance. La mise en marche de la régression OPG avec $\theta = \hat{\theta}$ fournit, alors, une manière simple de tester comment, en fait, sont satisfaites les CPO par une série d'estimations calculées au moyen d'un certain programme informatique.

Comme les estimations \hat{c} de la régression (2.24) valent zéro lorsque les régresseurs sont \hat{G} , ces régresseurs n'ont aucun pouvoir explicatif sur ι , et la

somme des résidus au carré (SSR) est alors égale au total de la somme des carrés.

Cette dernière vaut

$$i^T i = \sum_{i=1}^n 1 = n,$$

et l'estimation MV de la variance des résidus dans (2.24) est juste l'unité :

$$\frac{1}{n} SSR = \frac{1}{n} i^T i = \frac{1}{n} n = 1.$$

L'estimation de la matrice de covariance pour le vecteur \hat{c} de (2.24) est alors

$$(\hat{G}^T \hat{G})^{-1}.$$

C'est cette expression qui donne à la régression OPG son nom.

L'estimateur OPG a été préconisé par Brendt, Hall, Hall et Hausman (1974) et on s'y réfère sous le nom de l'estimateur BHHH.

Afin de nous aider à mieux comprendre les résultats théoriques énumérés ci-dessus, sur la méthode d'estimation MV, voyons un exemple. Cet exercice présente un intérêt. Tout d'abord, il fournit une illustration concrète de la manière d'utiliser la méthode MV. Deuxièmement, il fournit une matrice de covariance asymptotique pour les estimations de β et σ conjointement.

Considérons un modèle linéaire classique

$$y = X\beta + u \quad u \approx N(0, \sigma^2 I). \quad (2.25)$$

Dans ce modèle, la matrice des variables explicatives, X , est supposée être exogène. Le vecteur paramétrique β est supposé être de longueur k , ce qui implique qu'il y a $k+1$ paramètres à estimer, dont les k premiers sont les éléments du vecteur β , et le dernier est celui de σ . Les éléments u_i du vecteur u sont *iid* $\approx N(0, \sigma^2)$, et y_i est distribué, conditionnellement à X , selon $N(X_i \beta, \sigma^2)$.

La fonction de densité de u_t est

$$f(u_t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{u_t^2}{2\sigma^2}\right).$$

Afin de construire la fonction de vraisemblance, nous avons besoin de la fonction de densité de y_t plutôt que celle de u_t . La fonction qui relie u_t à y_t est

$$u_t = y_t - X_t\beta \quad (2.26)$$

La FDC de y_t s'écrit ainsi,

$$f_t(y_t, \beta, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_t - X_t\beta)^2}{2\sigma^2}\right). \quad (2.27)$$

La contribution à la fonction de logvraisemblance apportée par la $t^{\text{ième}}$ observation est le logarithme de (2.27). Comme $\log \sigma = \frac{1}{2} \log \sigma^2$, nous obtenons

$$l_t(y_t, \beta, \sigma) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y_t - X_t\beta)^2.$$

Comme toutes les observations sont indépendantes, la fonction logvraisemblance elle-même correspond précisément à la somme des contributions $l_t(y_t, \beta, \sigma)$ sur tout t , ou

$$\begin{aligned} l(y, \beta, \sigma) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - X_t\beta)^2. \\ &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)^T (y - X\beta). \end{aligned} \quad (2.28)$$

Pour trouver l'estimateur MV, nous devons maximiser (2.28) par rapport aux paramètres inconnus β et σ .

Pour le modèle de régression (2.25), le vecteur paramétrique θ est le vecteur $[\beta; \sigma]$. Nous calculons à présent la matrice d'information asymptotique $J(\beta, \sigma)$ en

utilisant la méthode basée sur la matrice CG, qui ne nécessite que les dérivées premières.

La dérivée première de $l_i(y_i, \beta, \sigma)$ par rapport à β_i est

$$G_{ii}(\beta, \sigma) = \frac{\partial l_i}{\partial \beta_i} = \frac{1}{\sigma^2} (y_i - X_i \beta) X_{ii} = \frac{1}{\sigma^2} u_i X_{ii}, \quad i = 1, \dots, k. \quad (2.29)$$

La dérivée première $l_i(y_i, \beta, \sigma)$ par rapport à σ est

$$G_{i,k+1}(\beta, \sigma) = \frac{\partial l_i}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1}{\sigma^3} (y_i - X_i \beta)^2 = -\frac{1}{\sigma} + \frac{1}{\sigma^3} u_i^2 \quad (2.30)$$

Les expressions (2.29) et (2.30) sont tout ce dont nous avons besoin pour calculer la matrice d'information en utilisant la matrice CG. La colonne de cette matrice qui correspond σ aura l'élément type (2.30), tandis que les k colonnes restantes, qui correspondent aux β_i , auront l'élément type (2.29).

Pour $i, j = 1, \dots, k$, le $ij^{ième}$ élément de $G^T G$ est

$$\sum_{t=1}^n \frac{1}{\sigma^4} u_t^2 X_{ti} X_{tj}.$$

Ainsi, l'élément $J(\beta, \sigma)$ correspondant à β_i et β_j est

$$J(\beta_i, \beta_j) = p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{t=1}^n \frac{u_t^2}{\sigma^4} X_{ti} X_{tj} \right).$$

Comme u_t^2 a une espérance de σ^2 sous le PGD caractérisé par (β, σ) et est indépendant de X , nous pouvons le remplacer ici par σ^2 pour obtenir

$$J(\beta_i, \beta_j) = p \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{t=1}^n \frac{1}{\sigma^2} X_{ti} X_{tj} \right).$$

Ainsi, nous voyons que le bloc entier (β, β) de la matrice d'information asymptotique est

$$\frac{1}{\sigma^2} p\lim_{n \rightarrow \infty} \left(\frac{1}{n} X^T X \right). \quad (2.31)$$

L'élément de $J(\beta, \sigma)$ correspondant à σ est

$$\begin{aligned} J(\sigma, \sigma) &= p\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\sigma^2} + \frac{u_i^4}{\sigma^6} - \frac{2u_i^2}{\sigma^4} \right) \right) \\ &= \frac{1}{n} \left(\frac{n}{\sigma^2} + \frac{3n\sigma^4}{\sigma^6} - \frac{2n\sigma^2}{\sigma^4} \right) \\ &= \frac{2}{\sigma^2}. \end{aligned} \quad (2.32)$$

Ici, nous avons utilisé les faits que, sous le PGD caractérisé par (β, σ) , $E(u_i^2) = \sigma^2$ et $E(u_i^4) = 3\sigma^4$, la dernière égalité étant une propriété bien connue de la distribution normale (voir chapitre 1).

Le $(i, k+1)^{ième}$ élément de $(G^T G)$ est

$$\begin{aligned} &\sum_{i=1}^n \left(-\frac{1}{\sigma} + \frac{1}{\sigma^3} u_i^2 \right) \left(\frac{1}{\sigma^2} u_i X_{ii} \right) \\ &= -\sum_{i=1}^n \frac{1}{\sigma^3} u_i X_{ii} + \sum_{i=1}^n \frac{1}{\sigma^5} u_i^3 X_{ii} \end{aligned} \quad (2.33)$$

Finalement, l'élément de $J(\beta, \sigma)$ correspondant à β_i et σ est

$$\begin{aligned} J(\beta_i, \sigma) &= p\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \left(-\frac{u_i X_{ii}}{\sigma^3} + \frac{u_i^3 X_{ii}}{\sigma^5} \right) \right) \\ &= 0. \end{aligned} \quad (2.34)$$

Les éléments sont nuls parce que, sous le PGD caractérisé par (β, σ) et u_i est indépendant de X , et le fait que les aléas soient normalement distribués implique que $E(u_i) = E(u_i^3) = 0$.

En regroupant les résultats (2.31), 2.32) et (2.34), nous obtenons

$$J(\beta, \sigma) = \frac{1}{\sigma^2} \begin{bmatrix} p\lim(n^{-1} X^T X) & 0 \\ 0^T & 2 \end{bmatrix}. \quad (2.35)$$

Dans la pratique, naturellement, nous sommes intéressés par $\hat{\beta}$ et $\hat{\sigma}$. Ainsi, au lieu d'utiliser (2.35), nous devrions en fait réaliser des inférences basées sur la matrice de covariance estimée. Cette dernière est obtenue en multipliant l'inverse de (2.35) par n^{-1} et en remplaçant σ par $\hat{\sigma}$. D'où

$$\hat{Var}(\hat{\theta}) = \hat{Var}(\hat{\beta}, \hat{\sigma}) = \begin{bmatrix} \hat{\sigma}^2 (X^T X)^{-1} & 0 \\ 0^T & \hat{\sigma}^2 / 2n \end{bmatrix} \quad (2.36)$$

2.2.2 Estimation des modèles ARCH et GARCH

Dans le chapitre 1, nous avons vu que l'équation de la variance conditionnelle (fonction de variance) d'un modèle GARCH(1,1) dépend des carré des résidus retardés, qui à leur tour dépendent des estimations de la fonction de régression. Pour cette raison, l'estimation de la fonction de régression ainsi que la fonction de variance doivent être faites ensemble. L'estimation simultanée ces deux fonctions en même temps s'avère plus efficace qu'estimer l'une conditionnellement à l'autre, comme c'est le cas lorsqu'on utilise les moindres carrés généralisés (MCG); voir Engle (1982).

La manière recommandée, pour estimer les modèles avec erreurs GARCH est de supposer que les termes d'erreur sont normalement distribués et d'estimer ensuite les fonctions du modèle par MV.

Supposons un modèle de régression linéaire, où les erreurs suivent un processus GARCH et les innovations sont normalement distribuées

$$y_t = X_t \beta + u_t,$$

$$\text{et } u_t = \sigma_t \varepsilon_t; \quad \sigma_t^2 \equiv E(u_t^2 | \Omega_{t-1}) = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \delta_j \sigma_{t-j}^2 \quad (2.37)$$

où $\varepsilon_t \approx N(0,1)$, et Ω_{t-1} dénote l'ensemble de l'information disponible jusqu'à $t-1$.

L'équation de la régression linéaire peut se réécrire, comme

$$\frac{y_t - X_t \beta}{\sigma_t(\beta, \theta)} = \varepsilon_t, \quad (2.38)$$

où y_t est la variable dépendante, X_t est un vecteur d'exogène ou de régresseurs prédéterminés, et β est un vecteur des paramètres de la régression. La fonction $\sigma_t^2(\beta, \theta)$ est définie pour un choix particulier de retards p et q par l'équation (2.37) où u_t est remplacé par $y_t - X_t \beta$. Elle dépend donc de β mais aussi de α_i et δ_j que nous avons regroupé dans un seul vecteur, θ .

La densité de y_t conditionnelle à Ω_{t-1} s'écrit alors

$$\frac{1}{\sigma_t(\beta, \theta)} \phi\left(\frac{y_t - X_t \beta}{\sigma_t(\beta, \theta)}\right), \quad (2.39)$$

où $\phi(\bullet)$ dénote la densité normale standard. Et le premier facteur de (2.39) est le facteur Jacobien qui reflète le fait que la dérivée de ε_t par rapport à y_t vaut $\sigma_t^{-1}(\beta, \theta)$; voir Davidson et MacKinnon (1999).

En prenant le logarithme de l'équation (2.39), nous obtenons la contribution à la fonction logvraisemblance faite par la $t^{\text{ième}}$ observation, c'est-à-dire

$$l_t(\beta, \theta) = \frac{1}{2} \log 2\pi - \frac{1}{2} \log(\sigma_t^2(\beta, \theta)) - \frac{1}{2} \frac{(y_t - X_t \beta)^2}{\sigma_t^2(\beta, \theta)}. \quad (2.40)$$

Malheureusement, (2.40) ne peut pas être évaluée directement, à cause de la fonction skédastique $\sigma_t^2(\beta, \theta)$ qui est définie implicitement par l'équation réursive (2.37). Cette récursion ne constitue pas une définition complète parce qu'elle ne fournit pas les valeurs de départ requises pour débiter la récursion. En cherchant

les valeurs de départ appropriées, nous faisons face à la difficulté qu'il n'existe pas de solution analytique pour la densité d'un GARCH stationnaire.

Dans le cas d'un ARCH(q), nous pouvons éviter ce problème en traitant les q premières observations comme valeurs initiales, car dans ce cas, la fonction de variance est complètement déterminée par les q décalages des carrés des résidus. Il n'y a donc pas d'information manquante pour les $q + 1$ observations jusqu'à n . Nous pouvons donc sommer les contributions (2.40), juste pour ces observations.

Mais une telle approche fonctionne uniquement avec des modèles ARCH, et dans la pratique ces modèles sont moins employés dans la modélisation des rendements financiers.

Avec un modèle GARCH(p, q), les p valeurs de départ de σ_i^2 sont requises en plus des $q - p$ valeurs de départ des carrés des résidus pour pouvoir débiter la récursion (2.37). Il est donc nécessaire de recourir à une sorte de procédé ad hoc pour spécifier les valeurs de départ. Une possibilité non intéressante serait de les poser à zéro. Une meilleure idée est de les remplacer par une estimation moyenne. Pour ce faire, il existe au moins deux manières différentes de procéder.

La première consiste à remplacer l'espérance non conditionnelle par la fonction des paramètres θ ; ceci conduit à prendre $\left(\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \delta_1} \right)$ dans le cas d'un GARCH(1,1).

La seconde consiste à utiliser la somme des carrés des résidus des estimations MCO divisé par n .

Une autre approche est de considérer les valeurs de départ inconnues comme des paramètres supplémentaires, et de maximiser la fonction logvraisemblance par

rapport à ces paramètres β et θ conjointement. Dans tous les échantillons, sauf pour des grands échantillons, le choix des valeurs de départ peut avoir des effets significatifs sur les estimations des paramètres. Ceci implique que l'utilisation de programmes différents pour l'estimation des GARCH peut produire des résultats très différents. Brooks et Persaud (2001) ont développé une argumentation convaincante sur cette problématique.

Peu importe le choix des valeurs de départ, la maximisation de la fonction de logvraisemblance obtenue par l'addition des contributions n'est pas chose facile, surtout lorsqu'il s'agit de modèles GARCH.

Il est essentiel d'utiliser, autant que possible, les dérivées analytiques, plutôt que numériques, voir à ce sujet Fiorentini, Calzolari et Panattoni (1996).

Par ailleurs, selon McCullough et Renfro (1999) et Brooks, Burke et Persaud (2001) tous les logiciels ne fournissent pas des estimations et des écarts types fiables. Par conséquent, il est vivement recommandé d'estimer ce type de modèle plus d'une fois en utilisant des options différentes et des programmes différents.

2.2.3 Les dérivées de la log-vraisemblance d'un modèle ARCH(1)

Dans la sous-section précédente, nous nous sommes aperçus que l'estimation par MV requiert des calculs de dérivées afin de déterminer le(s) estimateur(s) qui maximise(nt) la fonction de vraisemblance ou la fonction de logvraisemblance. Ainsi, dans la présente sous-partie et la suivante, nous allons calculer les dérivées nécessaires à l'estimation par MV des modèles ARCH-GARCH.

Supposons le modèle de régression linéaire :

$$y_t = X_t \beta + u_t \quad (2.41)$$

nous supposons que les innovations suivent un processus ARCH(1) :

$$u_t = \sigma_t v_t; \sigma_t^2 = \alpha + \gamma u_{t-1}^2$$

où v_t est $BB \approx N(0,1)$.

Nous avons vu que la contribution de l'observation t à la fonction de la logvraisemblance a la forme

$$l_t(\alpha, \beta, \gamma) = -\frac{1}{2} \log(2\pi\sigma_t^2) - \frac{1}{2\sigma_t^2} u_t^2(\beta) ;$$

où $u_t(\beta) = y_t - X_t \beta$

Nous pouvons dès lors déterminer les dérivées analytiques de l_t par rapport aux différents paramètres β, α et γ .

Nous avons tout d'abord :

$$\frac{\partial l_t}{\partial u_t} = \frac{u_t}{\sigma_t^2}; \frac{\partial l_t}{\partial \sigma_t^2} = \frac{u_t^2 \sigma_t^2}{\sigma_t^4}$$

Par la suite, nous avons :

$$\frac{\partial \sigma_t^2}{\partial u_{t-1}} = 2\gamma u_{t-1}$$

Ainsi, par rapport aux paramètres, nous avons donc

$$\frac{\partial u_t}{\partial \beta} = -X_t; \frac{\partial u_t}{\partial \alpha} = 0; \frac{\partial u_t}{\partial \gamma} = 0$$

$$\frac{\partial \sigma_t^2}{\partial \beta} = -2\gamma u_{t-1} X_{t-1}; \frac{\partial \sigma_t^2}{\partial \alpha} = 1; \frac{\partial \sigma_t^2}{\partial \gamma} = u_{t-1}^2$$

Il vient alors :

$$\frac{\partial l_t}{\partial \beta} = \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \beta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta} = \frac{u_t X_t}{\sigma_t^2} - 2\gamma \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} u_{t-1} X_{t-1}$$

$$\frac{\partial l_t}{\partial \alpha} = \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2 \sigma_t^4}$$

$$\frac{\partial l_t}{\partial \gamma} = \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \gamma} = \frac{u_t^2 - \sigma_t^2}{2 \sigma_t^4} u_{t-1}^2$$

Ces dérivées peuvent se généraliser dans le cas d'un ARCH(q). Toutefois, leurs calculs demeurent fastidieux.

2.2.4 Les dérivées de la log-vraisemblance d'un modèle GARCH(1,1)

On maintient le modèle (2.41), mais on suppose que les aléas du modèle suivent un processus GARCH(1,1).

La caractérisation formelle du processus est alors :

$$u_t = \sigma_t v_t; \sigma_t^2 = \alpha + \gamma u_{t-1}^2 + \delta \sigma_{t-1}^2 \quad (2.42)$$

où v_t est BB normal à variance 1.

La contribution à la fonction de log-vraisemblance est toujours de la forme

$$l_t(\alpha, \beta, \gamma, \delta) = -\frac{1}{2} \log(2\pi\sigma_t^2) - \frac{1}{2\sigma_t^2} u_t^2(\beta);$$

où $u_t(\beta) = y_t - X_t \beta$

Afin de rendre le calcul des dérivées plus sympathique, on va supposer que la première observation porte l'indice 0 plutôt que 1, et nous allons définir la convolution de deux séries.

Définition : La convolution de deux séries x_t et $y_t, t = 0, 1, \dots, n$ est définie par

$$conv(x, y)_t = \sum_{s=0}^t x_s y_{t-s}.$$

On a alors :

$$\frac{\partial l_t}{\partial u_t} = \frac{u_t}{\sigma_t^2}; \frac{\partial l_t}{\partial \sigma_t^2} = \frac{u_t^2 - \sigma_t^2}{2 \sigma_t^4}$$

$$\frac{\partial u_t}{\partial \beta} = -X_t; \frac{\partial u_t}{\partial \alpha} = 0; \frac{\partial u_t}{\partial \gamma} = 0; \frac{\partial u_t}{\partial \delta} = 0$$

Nous constatons que les dérivées $\frac{\partial l_t}{\partial u_t}$ et $\frac{\partial l_t}{\partial \sigma_t^2}$ sont inchangées par rapport au processus ARCH(1)=GARCH(1,0), de même que les dérivées de u_t par rapport aux paramètres. Par contre, il faut déterminer à nouveau les dérivées de σ_t^2 .

Pour ce faire, nous réécrivons, après manipulation de (2.42), l'équation (2.42) afin d'éliminer le terme constant α . Pour ce faire, on pose

$$S_t^2 = \sigma_t^2 - \alpha_{(1-\delta)}, \text{ et nous obtenons}$$

$$S_t^2 = \delta S_{t-1}^2 + \gamma u_{t-1}^2$$

qui est une équation de récurrence.

Cette récurrence conduit aux équations suivantes :

$$S_2^2 = \delta S_1^2 + \gamma u_1^2, S_3^2 = \delta^2 S_1^2 + \delta \gamma u_1^2 + \gamma u_2^2$$

$$S_4^2 = \delta^3 S_1^2 + \delta^2 \gamma u_1^2 + \delta \gamma u_2^2 + \gamma u_3^2 \text{ et ainsi de suite.}$$

L'expression de S_t^2 pour tout t est donc $S_t^2 = \delta^{t-1} S_1^2 + \gamma \sum_{r=0}^{t-2} \delta^r u_{t-r-1}^2$

ou en utilisant la définition de la convolution, c'est-à-dire :

$$S_t^2 = \delta^{t-1} S_1^2 + conv(\delta, u^2)_{t-1} \quad (2.43)$$

où nous notons u^2 le vecteur dont l'élément est u_t^2 , et δ le vecteur dont l'élément t est δ^{t-1} .

Nous pouvons donc ainsi calculer la variance non conditionnelle du processus GARCH(1,1) en utilisant la récurrence dans l'équation (2.42).

En notant cette variance σ_0^2 , nous remarquons que les espérances marginales de σ_t^2 et de u_t^2 sont égales à σ_0^2 nous avons alors : $\sigma_0^2 = \alpha + \gamma\sigma_0^2 + \delta\sigma_0^2$, d'où

$$\sigma_0^2 = \frac{\alpha}{1 - \gamma - \delta}.$$

Par conséquent,

$$s_0^2 = \sigma_0^2 - \frac{\alpha}{1 - \delta} = \frac{\alpha\delta}{(1 - \gamma - \delta)(1 - \delta)}.$$

Pour initialiser la récurrence et rendre opérationnelle l'équation (2.43), nous supposons que $s_1^2 = s_0^2$.

Il vient alors

$$S_t^2 = \frac{\alpha\gamma\delta^{t-1}}{(1 - \gamma - \delta)(1 - \delta)} + \text{conv}(\delta, \mu^2)_{t-1} \text{ et}$$

$$\sigma_t^2 = \frac{\alpha}{1 - \delta} + \frac{\alpha\gamma\delta^{t-1}}{(1 - \gamma - \delta)(1 - \delta)} + \text{conv}(\delta, u^2)_{t-1}$$

Ainsi, pour tout $t=1$, la valeur de la convolution est nulle. Dès lors, les dérivées nécessaires à l'évaluation de la logvraisemblance sont déterminées comme suit

$$\frac{\partial \sigma_t^2}{\partial \alpha} = \frac{1}{1 - \delta} + \frac{\gamma\delta^{t-1}}{(1 - \gamma - \delta)(1 - \delta)}$$

$$\frac{\partial \sigma_t^2}{\partial \gamma} = \frac{\alpha\delta^{t-1}}{(1 - \gamma - \delta)^2} + \text{conv}(\delta, u^2)_{t-1}$$

$$\frac{\partial \sigma_t^2}{\partial \delta} = \frac{\alpha}{(1 - \delta)^2} + \frac{\alpha\gamma\delta^{t-2}(t-1)}{(1 - \gamma - \delta)(1 - \delta)} + \frac{\alpha\gamma\delta^{t-1}(2 - \gamma - 2\delta)}{(1 - \gamma - \delta)^2(1 - \delta)^2} + \text{conv}(\delta, \frac{\partial}{\partial \delta} \delta, u^2)_{t-1}$$

$$\frac{\partial \sigma_t^2}{\partial \beta} = \text{conv}(\delta, \frac{\partial}{\partial \beta} u^2)_{t-1}$$

Étant donné que l'élément t du vecteur δ est δ^{t-1} , nous constatons que l'élément t du vecteur $(\partial/\partial\delta)\delta$ est $(t-1)\delta^{t-2}$. De même, l'élément t du vecteur $(\partial/\partial\beta)u^2$ est

$-2\mu_t X_t$. Nous aboutissons alors aux expressions

$$\frac{\partial l_t}{\partial \beta} = \frac{\partial l_t}{\partial u_t} \frac{\partial u_t}{\partial \beta} + \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta} = \frac{u_t X_t}{\sigma_t^2} - 2\gamma \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \text{conv}(\delta, u_t X_t)_{t-1}$$

$$\frac{\partial l_t}{\partial \alpha} = \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \alpha} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\frac{1}{1-\delta} + \frac{\gamma \delta^{t-1}}{(1-\gamma-\delta)(1-\delta)} \right)$$

$$\frac{\partial l_t}{\partial \gamma} = \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \gamma} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\frac{\alpha \delta^{t-1}}{(1-\gamma-\delta)^2} + \text{conv}(\delta, u_t^2)_{t-1} \right)$$

$$\frac{\partial l_t}{\partial \delta} = \frac{\partial l_t}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial \delta} = \frac{u_t^2 - \sigma_t^2}{2\sigma_t^4} \left(\frac{\alpha}{(1-\delta)^2} + \frac{\alpha \gamma \delta^{t-2(t-1)}}{(1-\gamma-\delta)(1-\delta)} + \frac{\alpha \gamma \delta^{t-1}(2-\gamma-2\delta)}{(1-\gamma-\delta)^2(1-\delta)^2} + \text{conv}((\partial/\partial\delta)\delta, u_t^2)_{t-1} \right)$$

Comme nous l'avons souligné précédemment, nous constatons bel et bien que la maximisation de la vraisemblance s'avère délicate, compte tenu de l'augmentation du nombre de dérivées à calculer. L'estimation des modèles ARCH-GARCH, requiert dans un premier temps la présence d'hétéroscédasticité conditionnelle dans les aléas du modèle spécifié. Pour ce faire, nous avons alors recours à un test ARCH, dont nous exposons la théorie dans la section suivante.

2.3 Test d'homoscédasticité : test ARCH

Considérons le modèle de régression linéaire avec erreur ARCH(q) défini par (2.41), c'est-à-dire :

$$y_t = X_t \beta + u_t$$

$$E(u_t | u_{t-1}) = 0$$

$$u_t = \sigma_t \nu_t, V(u_t | u_{t-1}) = \sigma_t^2 = h_t = \alpha + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + \dots + \gamma_q u_{t-q}^2$$

Sous l'hypothèse nulle d'erreurs homoscédastiques, tous les γ_i sont nuls.

Comme Engle (1982) l'a montré le premier, il est facile de tester cette hypothèse en exécutant la régression :

$$\hat{u}_t^2 = a + c_1 \hat{u}_{t-1}^2 + c_2 \hat{u}_{t-2}^2 + \dots + c_q \hat{u}_{t-q}^2 + \text{résidu}, \quad (2.44)$$

où \hat{u}_t désigne un résidu provenant de l'estimation par moindres carrés du modèle de régression auquel sont associés les u_t . Nous calculons ensuite un test F ordinaire (ou simplement n fois le R^2 centré) pour les hypothèses de nullité des paramètres c_1 à c_q . Il ne s'agit nul autre que d'une régression artificielle où les régresseurs sont des résidus au carré retardés. Ainsi, pour tout modèle de régression estimé à l'aide de séries temporelles, il est très facile de tester l'hypothèse nulle d'homoscédasticité contre l'hypothèse alternative que les erreurs suivent un processus ARCH(q). Il s'agit simplement d'un test LM pour l'hétéroscédasticité conditionnelle en termes de régression artificielle.

2.4 Tests d'asymétrie et d'aplatissement

Bien qu'il soit correct d'utiliser les MCO chaque fois que les aléas associés à une fonction de régression sont d'espérance nulle et ont une matrice de covariance qui satisfait des conditions de régularité assez faibles, les moindres carrés conduisent à un estimateur optimal seulement dans des circonstances particulières. Ainsi, l'information sur les seconds moments des aléas conduira, en général, à un gain d'efficacité dans l'estimation des paramètres de la fonction de régression. C'est aussi le cas pour les moments des aléas d'ordre supérieur. Par exemple, si les aléas sont très leptokurtiques, c'est-à-dire si leur distribution possède des queues très épaisses, les MCO peuvent se révéler très inefficaces par rapport à un autre estimateur qui prend en compte la leptokurticité. Par analogie, les aléas sont

asymétriques, il sera possible de trouver mieux que les MCO en utilisant un estimateur qui reconnaît la présence d'asymétrie.

Tout cela suggère qu'il est généralement prudent de tester l'hypothèse que les aléas sont normalement distribués. Dans la pratique, nous calculons rarement des moments au-delà du troisième ou du quatrième; ceci signifie que l'on teste l'asymétrie ou l'excès du kurtosis (aplatissement). Dans une section antérieure, nous avons vu que pour une distribution normale de variance σ^2 , le troisième moment centré, qui détermine l'asymétrie, est nul, tandis que le quatrième moment centré, qui détermine l'excès du kurtosis, est $3\sigma^4$. Si le troisième moment n'est pas nul, la distribution est dite asymétrique. Si le quatrième moment centré est supérieur à $3\sigma^4$, la distribution est dite leptokurtique alors que si le quatrième moment centré est inférieur à $3\sigma^4$, la distribution est dite platykurtique. Dans la pratique, les résidus sont fréquemment leptokurtiques et rarement platykurtiques.

On rencontre souvent l'asymétrie et l'excès de kurtosis dans les données de rendements financiers, et surtout lorsque ces rendements sont mesurés sur de courtes périodes. Un bon modèle devrait éliminer, ou au moins réduire sensiblement, l'asymétrie et l'excès de kurtosis qui sont généralement présents dans des données quotidiennes, hebdomadaires et, moindrement, dans les rendements mensuels.

Un moyen simple pour tester si les aléas sont normalement distribués est d'imaginer un test sur la régression OPG (voir Russell Davidson et James G. Mackinnon, 1999, chapitre 13).

Supposons que le modèle de régression linéaire à tester soit

$$y_i = X\beta + \mu, \quad u \approx N(0, \sigma^2 I), \quad (2.45)$$

l'ajout d'une constante ou l'équivalent dans les régresseurs ne modifie pas les résultats. La régression OPG correspondant à ce modèle s'écrit

$$l = \frac{1}{\sigma^2} u_i(\beta) X_i b + b_\sigma \frac{u_i^2(\beta) - \sigma^2}{\sigma^3} + \text{résidu} \quad (2.46)$$

où l , une fois de plus, est un vecteur unitaire et nous avons posé $u_i \equiv y_i - X_i \beta$, et sous l'hypothèse que les aléas sont normalement distribués, il y a ni asymétrie ni excès de kurtosis. Si nous voulons tester l'hypothèse nulle que les aléas sont symétriques, contre l'hypothèse alternative que les aléas sont asymétriques, le régresseur naturel approprié à ajouter à (2.46) est $u_i^3(\beta)$.

L'estimation par MCO de (2.45) nous fournit $\hat{\beta}$ et $\hat{\sigma}^2 \equiv SSR/n$. Ainsi, la régression OPG associée à cette équation s'écrit

$$l = \frac{1}{\hat{\sigma}^2} u_i(\hat{\beta}) X_i b + b_\sigma \frac{u_i^2(\hat{\beta}) - \hat{\sigma}^2}{\hat{\sigma}^3} + c u_i^3(\hat{\beta}) + \text{résidu} \quad (2.47)$$

La statistique de test sera simplement le t de Student associé à ce régresseur, c'est-à-dire qu'on teste si $c = 0$.

La régression (2.47) n'est pas nécessairement compliquée. Car sous l'hypothèse nulle, le régresseur de test est déjà asymptotiquement orthogonal au second régresseur (c'est celui qui correspond au paramètre σ). Pour s'en rendre compte, évaluons les régresseurs par rapport au vrai paramètre β_0 au lieu de $\hat{\beta}$. Les résidus $u_i(\beta_0)$ correspondent alors aux aléas u_i et nous constatons que

$$p \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{u_i^2 - \sigma^2}{\sigma^3} u_i^3 = 0$$

Ce résultat est obtenu en utilisant la loi des grands nombres et l'hypothèse que les aléas (u_i) sont normalement distribués, ce qui implique que les moments impairs

sont nuls ($E(u_i^5) = E(u_i^3) = 0$). Donc, même si nous omettons le régresseur correspondant à σ , dans (2.47), le t -test de $c = 0$ reste asymptotiquement inchangé.

Rappelons que cette statistique de test restera inchangée, même dans un échantillon fini, si nous ajoutons à $u_i^3(\beta)$ n'importe quelle combinaison linéaire de régresseurs correspondant à β . Nous faisons référence ici au théorème FWL. Comme nous avons supposé qu'il y a un terme constant dans la régression, la statistique de test restera inchangée si nous remplaçons $u_i^3(\beta)$ par $u_i^3(\beta) - 3\sigma^2 u_i(\beta)$. En procédant ainsi, ce nouveau régresseur est asymptotiquement orthogonal à tous les régresseurs correspondant β , comme nous pouvons le constater par le calcul suivant :

$$p \lim_{n \rightarrow \infty} \frac{1}{n} \sum X_i u_i (u_i^3 - 3\sigma^2 u_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i E(u_i^4 - 3\sigma^2 u_i^2) = 0.$$

Sous l'hypothèse de normalité de u_i , $E(u_i^4) = 3\sigma^4$. D'où le résultat de la deuxième égalité ci-dessus. Par conséquent, il n'y a aucune différence asymptotiquement si nous omettons les régresseurs qui correspondent à β .

Les arguments ci-dessus impliquent que nous pouvons obtenir un test valide simplement en employant le t -test de la régression suivante

$$l = c(u_i^3(\hat{\beta}) - 3\hat{\sigma}^2 u_i(\hat{\beta})) + \text{résidu}, \quad (2.48)$$

ce qui est numériquement identique au t -test de la moyenne d'un échantillon dont le modèle comprend un seul régresseur. Comme la $p \lim$ de la variance des erreurs vaut 1, et que les régresseurs et les régressés sont asymptotiquement orthogonaux, les deux t -tests sont alors asymptotiquement équivalents à

$$\frac{n^{-1/2} \sum_{i=1}^n (\hat{u}_i^3 - 3\hat{\sigma}_i^2 \hat{u}_i)}{(n^{-1} \sum_{i=1}^n (\hat{u}_i^3 - 3\hat{\sigma}_i^2 \hat{u}_i)^2)^{1/2}}, \quad (2.49)$$

où $\hat{u}_i \equiv u_i(\hat{\beta})$. Puisque les résidus de la régression par MCO comportent une somme constante égale à zéro et que sous $H_0 = \hat{\sigma}^2 = \sigma^2$, le numérateur de (2.49) vaut alors $n^{-1/2} \sum \hat{u}_i^3$. Le sixième moment de la distribution normale vaut alors $15\sigma^6$, et la p lim du dénominateur est égale à la racine carré de

$$E(u_i^6 - 6\sigma^2 u_i^4 + 9\sigma^4 u_i^2) = \sigma^6(15 - 18 + 9) = 6\sigma^6.$$

Sous l'hypothèse que les résidus sont normalisés, il vient alors, en remplaçant u_i par e_i et σ par 1, que la statistique de test pour l'asymétrie est

$$\tau_3 \equiv (6n)^{-1/2} \sum_{i=1}^n e_i^3 \quad (2.50)$$

Il s'en suit que la variance de $n^{-1/2} \sum \hat{u}_i^3$ vaut $6\sigma^6$. Alors que la variance de $n^{-1/2} \sum \hat{u}_i^3$, qui est égale à la variance de u_i^3 , vaut $15\sigma^6$. Nous constatons une inexactitude dans ces deux résultats, qui peut être due au fait que les résidus normalisés sont calculés à l'aide d'estimation de l'espérance et de l'écart type.

Par analogie, pour tester l'hypothèse que les aléas ont un quatrième moment égal à $3\sigma^4$, le régresseur naturel à ajouter à (2.46) est $u_i^3 - 3\sigma^4$. Il est facile de constater que ce nouveau régresseur est asymptotiquement orthogonal à tous les autres régresseurs de l'équation (2.46) et que le t -test reste inchangé même si nous ajoutons $6\sigma^5$ fois le régresseur correspondant à σ , nous obtenons alors $u_i^4 - 6\sigma^2 u_i^2 + 3\sigma^4$. Puis nous divisons cette expression par σ^4 , le t -test n'est toujours pas affecté, et nous obtenons la régression OPG correspondant à ce test

$$l = c(e_i^4 - 6e_i^2 + 3) + \text{résidu}, \quad (2.51)$$

sous l'hypothèse que les résidus sont normalisés. La statistique de test associée à ce test est asymptotiquement équivalente à

$$\tau_4 \equiv (24n)^{-1/2} \sum_{i=1}^n (e_i^4 - 3) \quad (2.52)$$

Tout comme τ_3, τ_4 est asymptotiquement distribuée selon $N(0,1)$ sous l'hypothèse nulle de normalité. Leurs carrés seront asymptotiquement distribués selon une $\chi^2(1)$. De plus, puisque ces deux statistiques sont indépendantes, la somme de leurs carrés sera asymptotiquement distribuée selon une $\chi^2(2)$. Nous écrivons alors

$$\tau_{3,4} = \tau_3^2, \tau_4^2. \quad (2.53)$$

La statistique (2.53) est connue sous l'appellation statistique Jarque et Bera (1980).

CHAPITRE 3

Application Empirique sur les rendements boursiers

3.1 Introduction

Nous avons vu dans les deux premiers chapitres les formulations des modèles avec hétéroscédasticité conditionnelle, la méthode d'estimation utilisée en pratique, ainsi que les procédures de test qui peuvent être associées à ces méthodes.

Dans ce chapitre, nous nous proposons d'illustrer certaines de ces notions en modélisant des séries de rendements boursiers, mensuels, de trois titres (IBM, General Electric et S&P500) et nous comparons les résultats des estimations faites avec ECTS, avec ceux obtenus en EVIEWS et RATS. Ainsi, nous obtenons des résultats en utilisant les dérivées analytiques et numériques du MV des modèles (G)ARCH.

Afin d'illustrer ce que nous avons vu au chapitre 2, à propos de l'estimation de ces modèles, nous porterons une attention particulière sur le ARCH(1) et le GARCH(1,1).

Les données proviennent de « The Center for Research in Security of the University of Chicago ». Chacune de ces séries mensuelles représente la moyenne, en fin de séance, des rendements de ses titres, durant les jours ouvrables correspondant au mois considéré. Ces séries comportent 888 observations.

Notons que cette analyse ne vise en aucun cas à vérifier une quelconque théorie financière.

3.2 La première partie : Les données et leurs propriétés statistiques

Dans notre application, nous étudierons d'abord ces différentes séries de données graphiquement (analyse descriptive). Par la suite nous nous interrogeons sur la question de la stationnarité des données, nous étudierons alors la stationnarité à l'aide de corrélogrammes, et nous la testerons en utilisant des tests appropriés. Puis, nous envisagerons un traitement (sous forme de transformation) de ces données afin de les stationnariser, s'il s'avère qu'elles ne le sont pas.

Enfin nous prouvons, successivement, par le calcul des différents Kurtosis, Skewness et les P-values du test de Jarque-Bera, que le processus est non gaussien et que sa variance est hétéroscédastique.

3.2.1 Approche descriptive

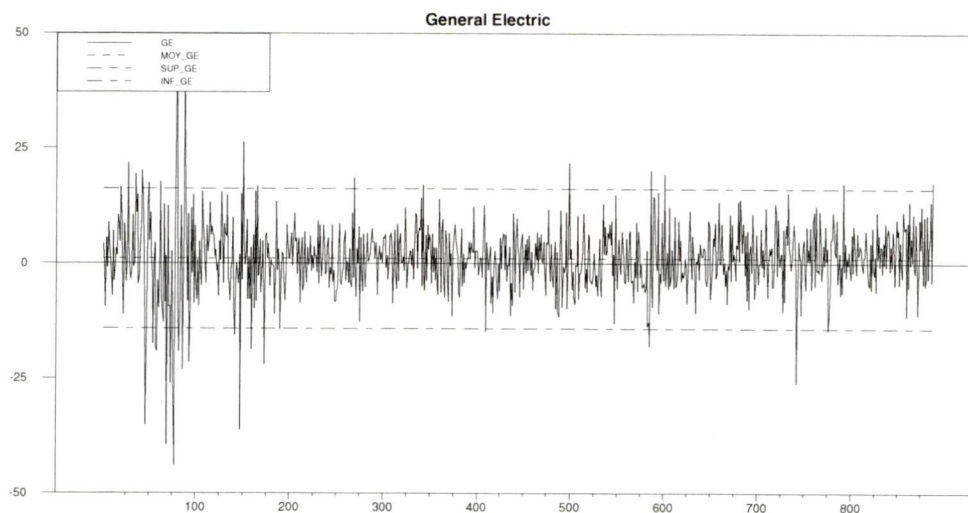
Une approche naturelle pour commencer l'étude d'un modèle est d'en faire une description approfondie afin d'avoir une idée de son comportement. Ceci passe avant tout par une approche descriptive du modèle.

Tout d'abord, il est fondamental de représenter les données dont nous disposons de façon à pouvoir analyser les comportements des séries. Les caractéristiques de ces séries nous permettent d'avoir une indication en particulier sur les valeurs extrêmes. Il est aussi possible de détecter les événements importants (chocs, changements de régime, stationnarité...) qui ont marqué la série financière en question.

La figure 3.1 représentant l'évolution temporelle des titres indique que ces séries semblent stationnaires : il n'existe pas de véritables tendances à la hausse et les moyennes des séries semblent se situer sur une droite parallèle à l'axe des abscisses. En effet, la moyenne, la variance et les covariances, de chacune des séries, semblent être relativement stables dans le temps. De plus, presque toutes les observations se trouvent à l'intérieur des bornes des intervalles de confiance de 95%. Notons, toutefois, certaines périodes plus volatiles que d'autres.

Le tableau 3.1 présente des statistiques sommaires des quatre séries. La pertinence de ces statistiques, à savoir : le coefficient d'aplatissement (Kurtosis) et le coefficient d'asymétrie (Skewness) sera analysée avec attention dans une section ultérieure.

Figure 3.1
Évolution temporelle des titres



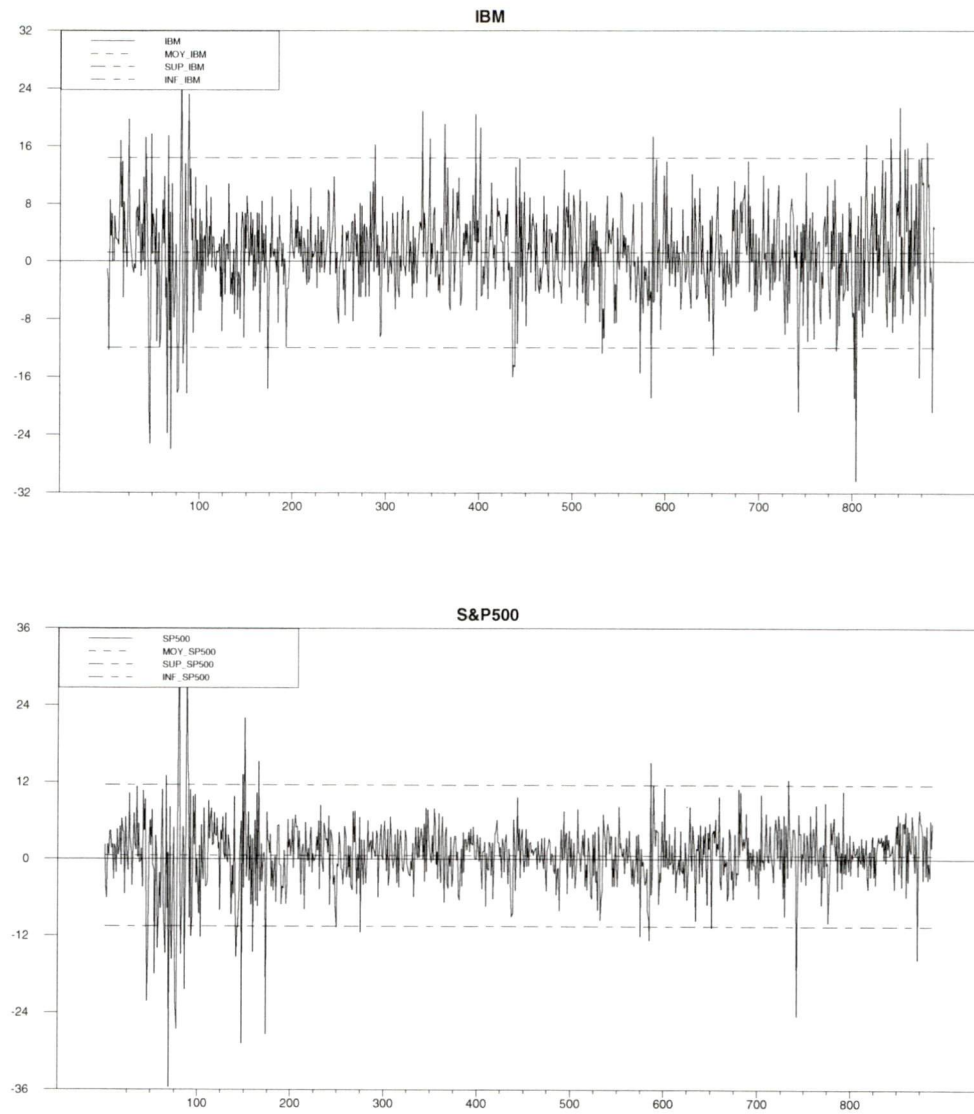
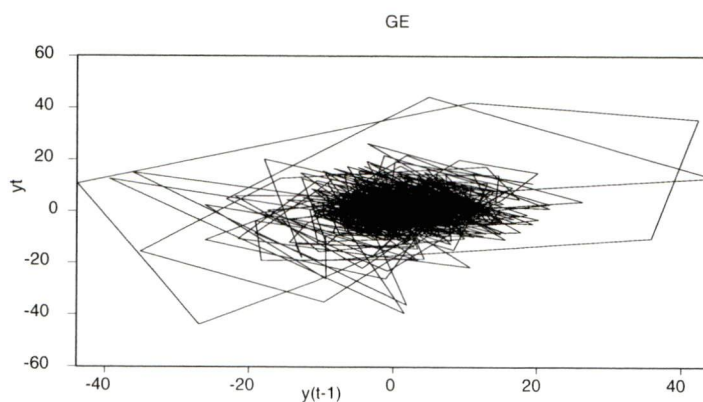


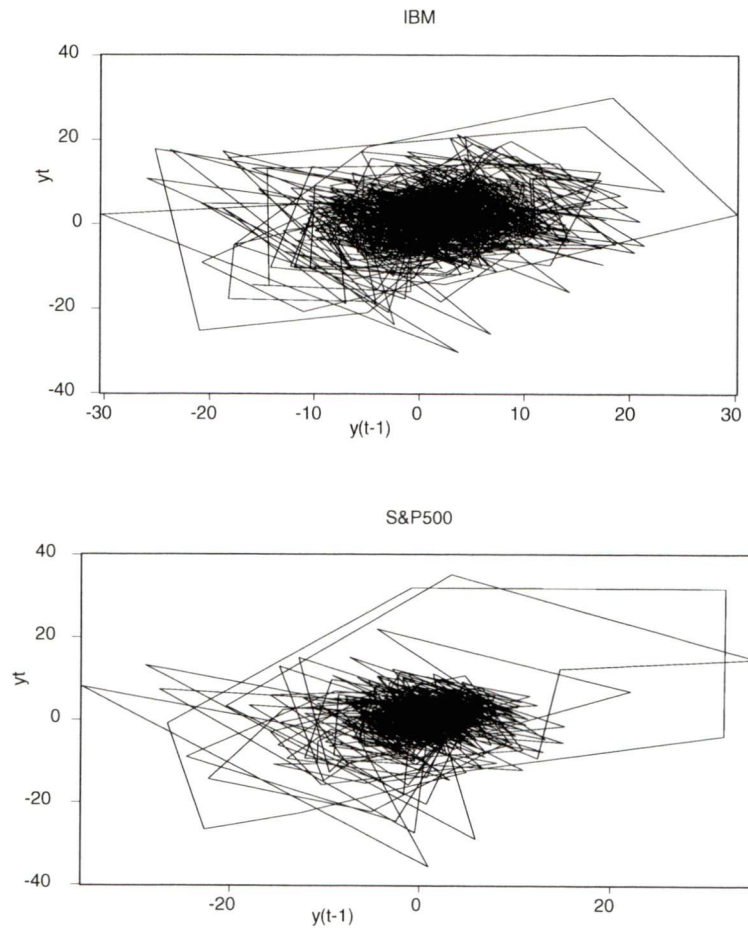
Tableau 3.1
Statistiques descriptives des séries mensuelles

	Mean	St-Error	Kurtosis	Skewness	Observations
GE	1.03757	7.74472	8.37465	-0.25208	888
IBM	1.24022	6.72867	4.92778	-0.23685	888
S&P500	0.53716	5.64471	11.14155	-0.52214	888

L'analyse descriptive des graphiques en nuage de points peut être révélateur du comportement des distributions des séries financières. Dans le chapitre 1, nous avons vu que les modèles (G)ARCH sont caractérisés par une distribution leptokurtique ou encore par des épisodes turbulents suivis d'épisodes plus calmes. Sur un graphique en nuage de point qui met en présence les valeurs courantes en fonction des valeurs retardées (figure 3.2), le caractère non gaussien des distributions devrait se traduire par une plus grande dispersion des différents couples $[Y, Y_{-i}]$. C'est ce qu'on observe sur les trois graphiques. De plus, le nuage semble plus compact mais touche une région plus étendue de façon dense : c'est exactement le comportement attendu pour un (G)ARCH avec des épisodes calmes (variance conditionnelle plus faible) et des épisodes plus turbulents (variance conditionnelle plus élevée).

Figure 3.2
 Y_t vs Y_{t-1}



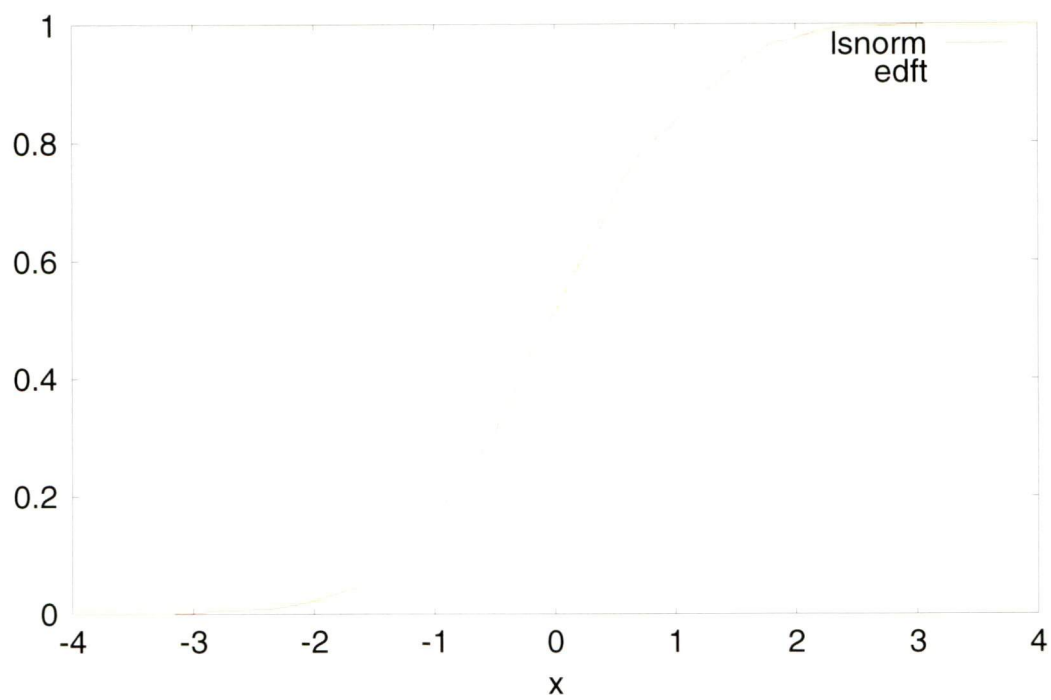


Par ailleurs, il est intéressant de comparer les répartitions empiriques des séries avec la loi de probabilité donnée par la théorie asymptotique. Cette loi est celle d'une normale.

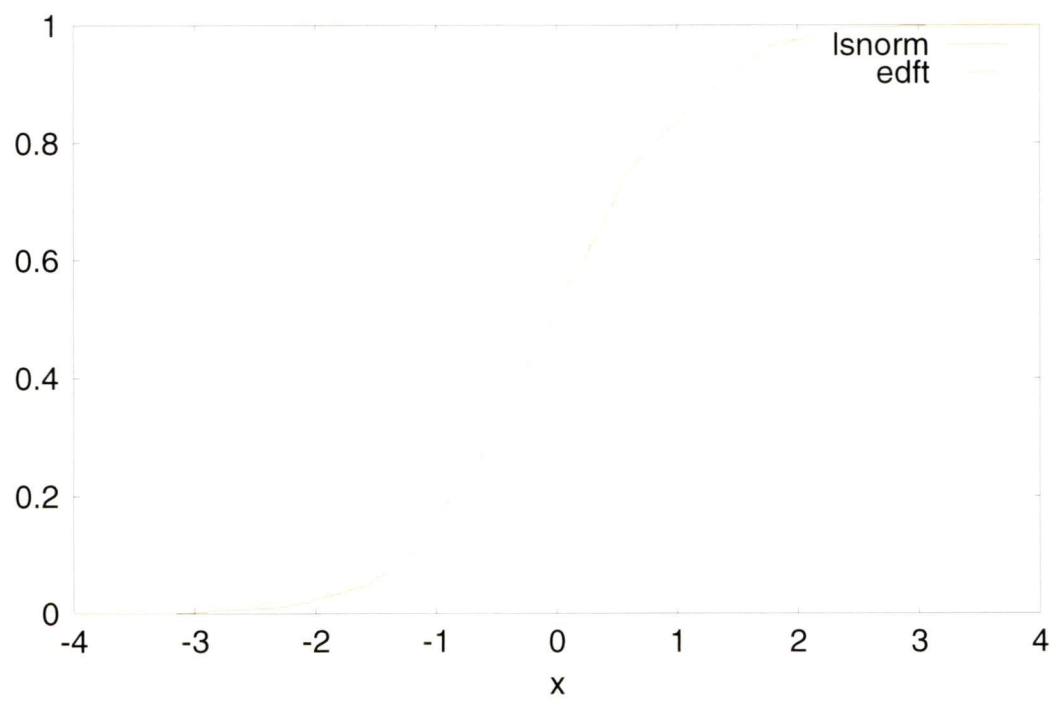
Nous voyons ces résultats dans la figure 3.3, où $edft$ et $lnorm$ représentent respectivement les fonctions théorique et empirique d'une série aléatoire centrée réduite et de séries financières. Il apparaît que la véritable loi de distribution des séries financières n'est pas identique à la loi normale, malgré l'hypothèse de normalité que l'on pose en estimant ces modèles.

En effet, on constate que ces dernières possèdent des queues plus épaisses que celle de la loi normale. Ce qui illustre le caractère leptokurtique de ces séries.

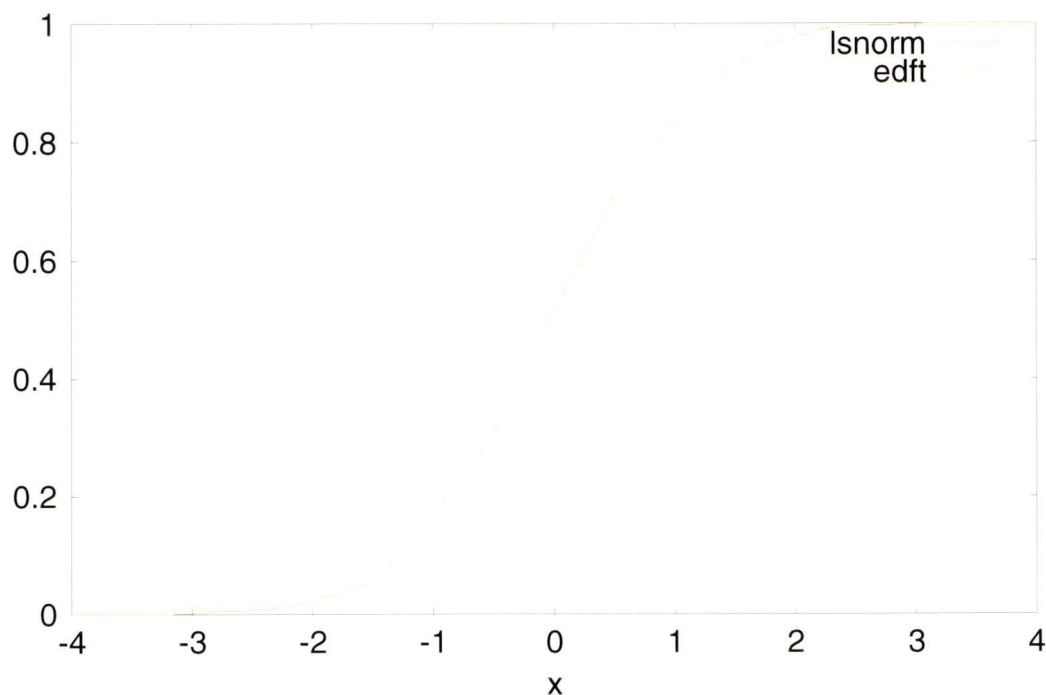
Figure 3.3
Fonctions de répartition empiriques et théoriques.
GE



IBM



S&P500



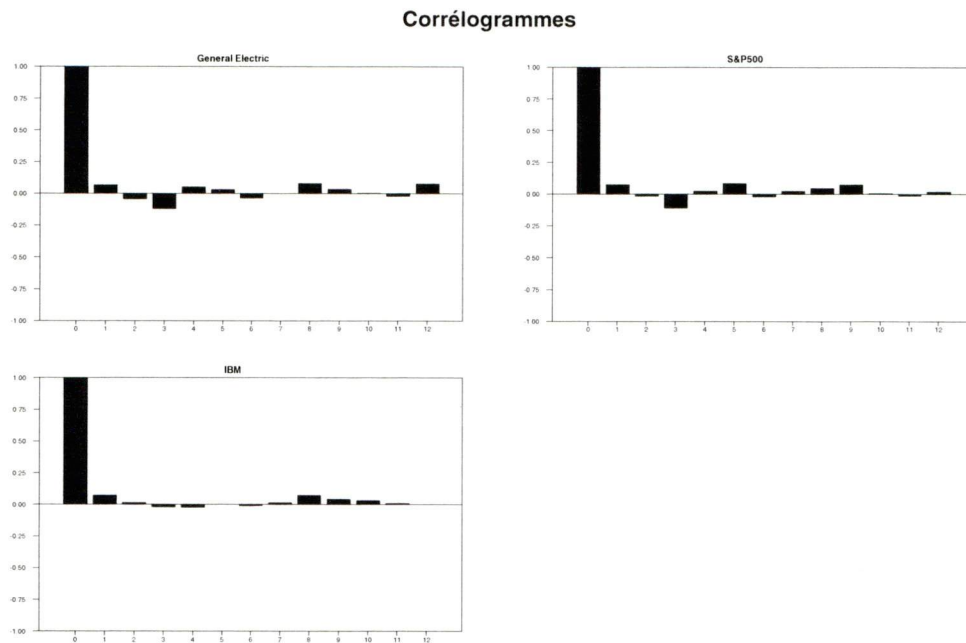
En examinant les corrélogrammes des séries (figure 3.4), nous constatons que ces corrélogrammes ne présentent pas de structure particulière : on n'observe pas d'autocorrélations persistantes pour les retards éloignés, un tel cas pourrait indiquer que l'hypothèse de non stationnarité est violée.

D'autre part, les séries ne sont pas bruit blanc. Notamment, les rendements de GE, de IBM et du S&P500 au temps t sont corrélés de façon significative aux rendements de la période précédente ($t-1$), respectivement. De plus, le test Ljung-Box nous indique que les autocorrélations sont conjointement différentes de 0. La statistique Q , associée à ce test, vaut $Q(24) = 52.96$, $Q(24) = 45.37$ et $Q(24) = 45.96$, pour GE, IBM et S&P500, respectivement.

Nous avons constaté que l'évolution des titres, à travers le temps (figure 3.1), semblait stationnaire et ce constat est appuyé par le corrélogramme. Il est donc

nécessaire de vérifier cette intuition en appliquant un test statistique de stationnarité.

Figure 3.4
Corrélogrammes des séries.



3.2.2 Test de stationnarité : Test de KPSS

Nous utilisons le test de stationnarité le KPSS (Kwiatkowski, Phillips, Schmidt et Shin) [1992]. Le choix de ce test est justifié par Lee et Schmidt (1996), lorsqu'ils ont montré que la puissance de ce test de stationnarité est considérable. Nous le préférons au test de la racine unitaire pour trois raisons qui sont :

- Le test de la racine unitaire n'est efficace que pour les séries linéaires.
- L'hypothèse nulle du test KPSS est la stationnarité alors qu'avec le test de la racine unitaire, nous testons la non stationnarité sous H_0 .
- La racine unitaire n'est qu'une parmi les formes de non stationnarité, c'est-à-dire, le fait de ne pas détecter une racine unitaire ne veut pas

dire impérativement que la série soit stationnaire, car il y a possibilité d'avoir d'autres formes de non stationnarité qui demeurent ombrées sans être identifiées.

La statistique du test KPSS pour la stationnarité est donnée par :

$$KPSS(q) = \frac{I}{n^2 \hat{\sigma}^2} \sum_{k=1}^n \left(\sum_{j=1}^k (Y_j - \bar{Y}_n) \right)^2$$

où $\hat{\sigma}^2(q)$ est l'estimateur HAC¹ des observations centrées $(Y_j - \bar{Y}_n)$.

L'estimateur (HAC) de la variance est donné par :

$$\hat{\sigma}^2(q) = \hat{\gamma}_0 + 2 \sum_{i=1}^q \omega_i \hat{\gamma}_i(q) \text{ avec } \omega_i(q) \equiv 1 - \frac{i}{q+1}$$

où :

$$\hat{\gamma}_0 = \frac{\sum_{t=1}^T \hat{u}_t^2}{T},$$

et $\hat{\gamma}_i$ est la simple auto-covariance au retard i .

Le test KPSS peut avoir deux sortes d'application. En effet, nous pouvons tester la stationnarité en niveau ($R_t = \mu + u_t$), comme nous pouvons aussi tester la stationnarité en tendance ($R_t = \mu + \beta t + u_t$). Il est constaté que la stationnarité en niveau est un cas particulier de la stationnarité en tendance (pour $t = 0$). Dans le cadre de notre application, nous avons appliqué le test de stationnarité KPSS sur les trois séries de rendements bousiers.

Dans ce test, on distingue ETA (μ) la statistique du test dans le cas de la stationnarité en niveau et ETA (τ) la statistique du test dans le cas de la stationnarité en tendance.

¹ Heteroskedastic and Autocorrelation Consistent : de Newey et West (1987)

- Application du KPSS sur les données

Tableau 3.2
Test de stationnarité.

Test KPSS	GE	IBM	S&P500
ETA(mu)	0.2473	0.1753	0.2889
ETA(tau)	0.0731	0.0834	0.0482
Valeur critique ETA(mu)	0.4630	0.4630	0.4630
Valeur critique ETA(tau)	0.1460	0.1460	0.1460

Ce tableau présente les valeurs obtenues de la statistique KPSS, pour les différents rendements boursiers de l'étude et en confrontant à la valeur critique du test, au seuil de 5 %, nous constatons que toutes les valeurs sont inférieures à la valeur critique. Donc l'hypothèse nulle de stationnarité est acceptée.

3.2.3 Test d'homoscédasticité

Tester la présence d'effet ARCH(q) consiste à faire une régression simple par les moindres carrés ordinaires de \hat{u}_t^2 sur $\hat{u}_t^2, \hat{u}_{t-1}^2, \dots$, et à considérer la statistique $LM=TR^2$, où R^2 et T sont respectivement coefficient de corrélation multiple et le nombre d'observations. Sous l'hypothèse nulle, cette statistique suit asymptotiquement une loi du Khi-deux avec autant de degrés de liberté que le nombre de retards associé à u_t . La régression a été faite pour plusieurs degrés de ARCH. Le tableau 3.2 présente la statistique de test, ainsi que les P-values correspondantes.

Tableau 3.3
Statistiques du test d'homoscédasticité.

	GE	IBM	S&P500
TR ²	171.7402	67.8958	179.4249
P-values	0.0000	0.0000	0.0000

Il est clair que toutes les P-values ne dépassent en aucun cas les 5 %, ce qui nous conduit à rejeter l'hypothèse nulle d'homoscédasticité, et à accepter l'hypothèse de l'hétéro-scédasticité. On corrobore ce test avec les corrélogrammes des résidus au carré.

Tableau 3.4
Corrélogrammes des résidus au carré.

GE(a) Sample: 1 888 Included observations: 888						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. **	. **	1	0.201	0.201	36.071	0.000
. **	. **	2	0.244	0.212	89.006	0.000
. **	. **	3	0.266	0.202	152.30	0.000
. **	. *	4	0.217	0.115	194.24	0.000
. *	. .	5	0.103	-0.033	203.74	0.000
. *	. .	6	0.153	0.039	224.74	0.000
. *	. *	7	0.190	0.103	257.11	0.000
. **	. **	8	0.264	0.198	319.60	0.000
. **	. *	9	0.247	0.147	374.51	0.000
. *	. .	10	0.163	-0.012	398.51	0.000

IBM(b)						
Sample: 1 888						
Included observations: 888						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *	. *	1	0.178	0.178	28.191	0.000
. *	. *	2	0.147	0.119	47.508	0.000
. *	. *	3	0.110	0.069	58.368	0.000
. *	. *	4	0.152	0.113	79.101	0.000
. *	. .	5	0.077	0.019	84.464	0.000
. *	. .	6	0.078	0.029	89.852	0.000
. *	. *	7	0.119	0.080	102.65	0.000
. *	. *	8	0.131	0.076	118.04	0.000
. .	. .	9	0.058	-0.006	121.11	0.000
. *	. .	10	0.088	0.039	128.02	0.000

S&P500(c)						
Sample: 1 888						
Included observations: 888						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. **	. **	1	0.216	0.216	41.651	0.000
. **	. *	2	0.213	0.174	81.953	0.000
. **	. **	3	0.280	0.221	151.80	0.000
. *	. .	4	0.159	0.049	174.37	0.000
. *	. .	5	0.102	-0.012	183.65	0.000
. *	. *	6	0.161	0.067	206.88	0.000
. *	. *	7	0.189	0.116	238.83	0.000
. **	. **	8	0.307	0.248	323.39	0.000
. **	. *	9	0.241	0.113	375.53	0.000
. **	. .	10	0.221	0.056	419.53	0.000

Dans les trois cas, la statistique Q le Ljung-Box, ainsi que la P-value qui lui est associée confirment le rejet de l'hypothèse d'homoscédasticité.

3.2.4 Test de normalité

L'analyse du skewness et du kurtosis (Tableau 3.1) nous a montré que l'hypothèse de normalité doit être rejetée, ce rejet provenant à la fois des moments d'ordre 3 et 4. Le fait d'adopter une représentation (G)ARCH, même avec des innovations

standardisées normales, permet d'obtenir des résidus dont la distribution non conditionnelle présente des queues plus épaisses que celle de la loi normale.

Rappelons que pour une loi normale centrée, le coefficient de skewness (τ_3 ou S) est nul et le coefficient de kurtosis (τ_4 ou K) vaut 3, cette loi est caractérisée par sa symétrie par rapport à la moyenne ainsi que par la faible probabilité des points extrêmes. Le test de normalité de Jarque et Bera (1980), que nous allons utiliser, est basé sur la définition des coefficients d'aplatissement et d'asymétrie (voir Chapitre II). Le tableau 3.5 présente les résultats de ce test.

Tableau 3.5
Résultats du test de normalité.

Normalité	GE	IBM	S&P500
Kurtosis	8.3746	4.9277	11.1415
Skewness	-0.2521	-0.2368	-0.5222
Jarque-Bera	1093.0881	148.3709	2524.9504
P-value	0.0000	0.0000	0.0000

En examinant les résultats, nous constatons :

- Un effet de leptokurtosis des résidus des différentes séries, car leurs valeurs de Kurtosis dépassent largement 3.
- Un effet d'asymétrie (ou biais) à gauche des distributions des résidus des différents titres, car leurs Skewness sont toutes négatives.
- Les différentes valeurs de la P-value du test de Jarque-Bera sont nulles, ce qui conduit à rejeter l'hypothèse de normalité des résidus.

Le rejet de l'hypothèse de normalité n'est pas imputable au caractère asymétrique de la distribution des résidus mais à un phénomène de leptokurticité qui affecte la distribution : les queues de cette distribution sont significativement plus épaisses

que celle qui résulteraient d'une distribution gaussienne dont les premiers moments seraient identiques.

Toutefois, malgré le fait que les résidus ne soient pas normaux, nous devons poser l'hypothèse qu'ils le sont en vue de construire et d'estimer les modèles (G)ARCH.

3.3 La deuxième partie : Estimation et spécification

Dans l'analyse descriptive des séries, nous avons constaté qu'aucune des trois séries ne semblait laisser se dégager une structure auto régressive (AR) ou moyenne mobile (MA) pertinente.

Par conséquent, le modèle de régression générale s'écrit comme suit :

$$R_t = \mu + u_t,$$

où $u_t = \varepsilon_t h_t$, ε_t iid, $E(\varepsilon_t) = 0$, $\text{var}(\varepsilon_t) = 1$ et h_t la fonction de variance conditionnelle.

Nous avons vu que la distribution des trois séries était leptokurtique. Les régressions linéaires effectuées sur les trois séries nous a conduit à une série de résidus qui ne suivent pas un bruit blanc d'après la statistique Q, qui indique la présence d'autocorrélation. Cette hypothèse est confirmée par l'étude de la fonction d'autocorrélation des séries u_t^2 , qui indiquent qu'il existe pour ces séries des autocorrélations significatives, même pour des retards assez importants.

Comme les cours du S&P500, du GE et de IBM sont corrélés et que leurs carrés le sont également, une possibilité alternative de modéliser ces séries est le processus

(G)ARCH. En effet, le test pour la présence d'effets ARCH (Tableau 3.3) indique clairement la présence d'un tel effet.

Dans la pratique, on suppose que ε_t suit une distribution normale ou une distribution t-Student.

3.3.1 Estimation de modèles (G)ARCH

Dans un premier temps, nous supposons que les erreurs sont de type ARCH(q). La fonction de variance conditionnelle h_t est de la forme de l'équation (1.9).

En estimant les modèles ARCH(q), $q = 1$, nous obtenons les résultats suivants :

$$GE_t = 1.1537 + u_t$$

$$\sigma_t^2 = 48.6588 + 0.1826u_{t-1}^2$$

$$(31.6915) \quad (6.0794)$$

$$IBM_t = 1.2552 + u_t$$

$$\sigma_t^2 = 37.4962 + 0.1678u_{t-1}^2$$

$$(22.5139) \quad (4.3098)$$

$$S\&P500_t = 0.7003 + u_t$$

$$\sigma_t^2 = 25.4897 + 0.1868u_{t-1}^2$$

$$(42.8841) \quad (5.2217)$$

où les nombres entre parenthèses sont les t-Student asymptotiques habituels.

Malgré que les coefficients de l'équation de la moyenne et ceux de l'équation de la variance étaient significatifs dans certains cas, aucune des séries ne peut être

modélisée en supposant que les erreurs sont de type ARCH(q). En effet, les tests de Ljung-Box sur les résidus et sur le carré des résidus ne nous ont pas permis de valider le modèle ARCH(1), car il y a autocorrélation entre les résidus au carré. Il aurait peut être fallu augmenter le nombre de paramètres q. Ce faisant, on aurait certainement violé les contraintes de positivité énumérée au Chapitre I. Et comme nous l'avons dit au premier Chapitre, un modèle GARCH(p,q) avec p et q petits peut approcher un modèle ARCH(q) où q est grand et modéliser les dynamiques de la variance sur de longue période.

Dans un deuxième temps, nous supposons que les erreurs sont de type GARCH(p,q), où p et q prennent les valeurs 1. La fonction de variance conditionnelle correspondante est de la forme de l'équation (1.11).

L'estimation du modèle GARCH(1,1) de la volatilité conditionnelle des différentes séries nous donne les résultats suivants :

$$\begin{aligned} GE_t &= 1.2334 + u_t \\ \sigma_t^2 &= 1.5841 + 0.8834\sigma_{t-1}^2 + 0.0972u_{t-1}^2 \\ &\quad (2.9355) \quad (41.4498) \quad (4.7152) \end{aligned}$$

$$\begin{aligned} IBM_t &= 1.3042 + u_t \\ \sigma_t^2 &= 3.0551 + 0.8345\sigma_{t-1}^2 + 0.1164u_{t-1}^2 \\ &\quad (3.3777) \quad (24.0488) \quad (4.3428) \end{aligned}$$

$$\begin{aligned} S\&P500_t &= 0.6859 + u_t \\ \sigma_t^2 &= 0.7636 + 0.8705\sigma_{t-1}^2 + 0.1119u_{t-1}^2 \\ &\quad (4.3323) \quad (62.0401) \quad (6.8136) \end{aligned}$$

où les nombres entre parenthèses sont les t-Student asymptotiques.

Au regard des résultats obtenus, le modèle GARCH(1,1) semble adéquat pour modéliser la volatilité conditionnelle des trois séries, car tous les coefficients de l'équation de la variance sont significatifs. Et les tests de Ljung-Box (Tableau 3.6) sur l'absence d'autocorrélation entre les résidus et entre les carrés des résidus nous permettent de ne pas rejeter l'hypothèse nulle et de valider ainsi le modèle.

Tableau 3.6
Test d'autocorrélation résiduelle.

	GE	IBM	S&P500
Q(10)	7.57(0.67)	15.92(0.10)	10.95(0.36)
Q(20) pour \hat{u}_t	16.94(0.65)	31.65(0.06)	24.00(0.24)
Q(10)	5.12(0.88)	2.80(0.98)	5.25(0.87)
Q(20) pour \hat{u}_t^2	15.65(0.73)	8.33(0.98)	9.68(0.97)

Dans le chapitre II, nous avons explicité la régression OPG. Et nous avons vu que, d'un point de vue pratique, cette régression (équation 2.24) permet aussi de vérifier les CPO. On note que si les ces conditions sont vérifiées, le R^2 et les paramètres estimés de cette régression, ainsi que les Students, sont égaux à zéro aux erreurs d'arrondi près. Les estimations effectuées, par MV, pour les trois modèles GARCH(1,1), vérifient bel et bien ces conditions. D'où les résultats :

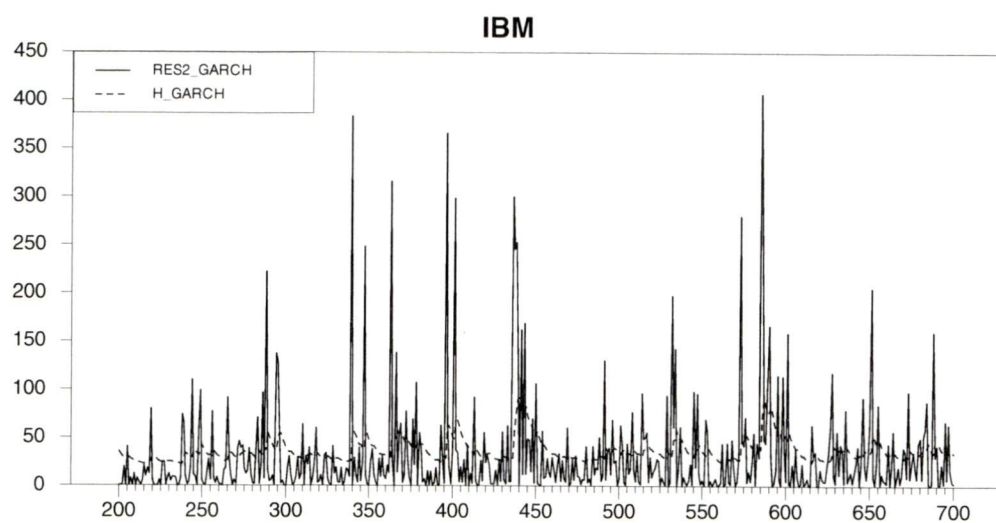
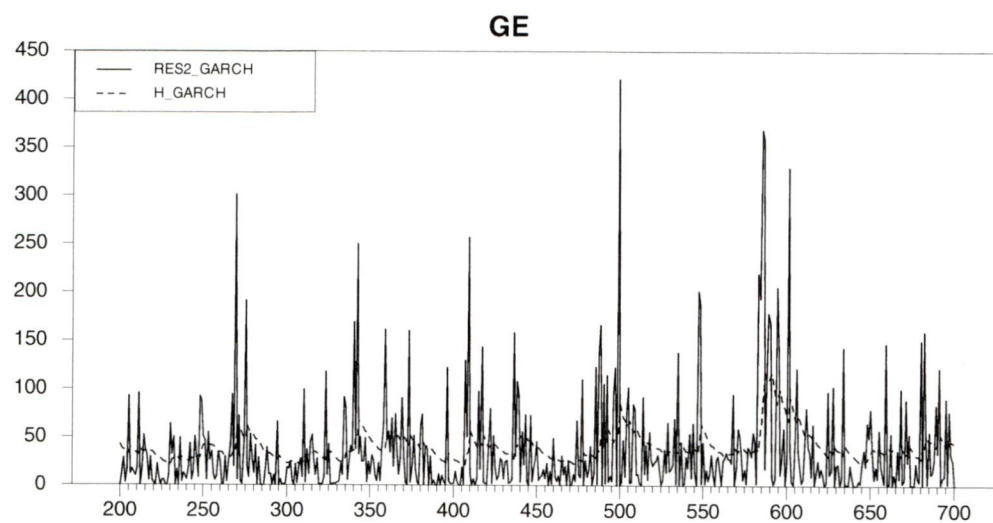
Tableau 3.7
Résultats de la régression OPG.

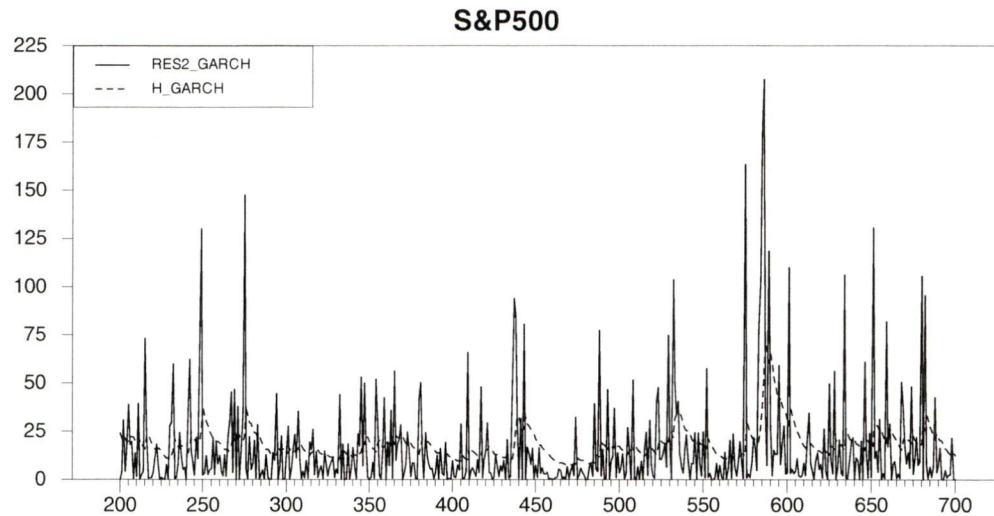
	GE	IBM	S&P500
CG1	-0.0046(-0.0085)	-1.08e-07(-1.19e-07)	-0.0006(-0.0038)
CG2	-0.0001(-0.0089)	-0.0000(-9.20e-08)	-0.0001(-0.0079)
CG3	0.0002(0.0108)	0.0000(9.66e-08)	0.0001(0.0085)
CG4	-0.0001(-0.0009)	-0.0000(-0.0000)	-0.0001(-0.0008)
R^2	1.4700e-07	0.0000	8.86e-08

Où les nombres entre parenthèse représentent t-Students habituels. Et les variables CG_i (Contribution au Gradient), $i=1, \dots, 4$, représentent respectivement les dérivées premières des paramètres $\alpha_0, \alpha_1, \delta_1$ et μ des équations de la variance conditionnelle et de la moyenne.

Lorsqu'on étudie les modèles (G)ARCH, l'hygiène économétrique nous oblige à analyser les résidus au carré, ainsi que la variance conditionnelle.

Figure 3.5
Résidus au carré et var.conditionnelle du GARCH(1,1).





En fait, l'idée derrière le graphique des résidus au carré (`res2_garch`) et de la variance conditionnelle (figure 3.5) nous permet de constater qu'en modélisant la variance de la série, à l'aide des modèles (G)ARCH, on élimine considérablement la volatilité, ainsi que les autocorrélations entre les erreurs. Car ces dernières sont à l'origine d'une persistance dans la variance. Nous obtenons alors une série plus lisse (`h_garch`).

L'analyse des corrélogrammes des carrés des résidus confirme que les résidus du modèle estimé ne présentent plus de corrélation sérielle et par conséquent, de l'hétéroscédasticité.

Tableau 3.8
Diagnostic de corrélation sérielle

GE(a)						
Sample: 1 888						
Included observations: 888						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	-0.030	-0.030	0.8065	0.369
. .	. .	2	0.001	0.000	0.8075	0.668
. .	. .	3	0.025	0.025	1.3513	0.717
. .	. .	4	-0.022	-0.020	1.7801	0.776
. .	. .	5	-0.043	-0.044	3.4269	0.634
. .	. .	6	0.001	-0.003	3.4273	0.754
. .	. .	7	-0.009	-0.007	3.4931	0.836
. .	. .	8	0.037	0.038	4.6950	0.790
. .	. .	9	0.010	0.011	4.7931	0.852
. .	. .	10	-0.019	-0.020	5.1272	0.883

IBM(b)						
Sample: 1 888						
Included observations: 888						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	0.020	0.020	0.3632	0.547
. .	. .	2	0.007	0.006	0.4031	0.817
. .	. .	3	-0.017	-0.017	0.6642	0.882
. .	. .	4	-0.013	-0.013	0.8200	0.936
. .	. .	5	0.013	0.013	0.9608	0.966
. .	. .	6	-0.032	-0.032	1.8624	0.932
. .	. .	7	-0.014	-0.013	2.0319	0.958
. .	. .	8	0.022	0.023	2.4621	0.963
. .	. .	9	0.001	0.000	2.4640	0.982
. .	. .	10	-0.020	-0.021	2.8087	0.986

S&P500(c)						
Sample: 1 888						
Included observations: 888						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1 -0.025	-0.025	0.5770	0.447	
. .	. .	2 -0.024	-0.025	1.1018	0.576	
. .	. .	3 0.007	0.006	1.1489	0.765	
. .	. .	4 -0.013	-0.013	1.2906	0.863	
. .	. .	5 -0.024	-0.024	1.7945	0.877	
. .	. .	6 0.008	0.006	1.8451	0.933	
. .	. .	7 -0.037	-0.037	3.0483	0.880	
. .	. .	8 0.038	0.037	4.3765	0.822	
. .	. .	9 0.031	0.031	5.2576	0.811	
. .	. .	10 -0.001	0.003	5.2585	0.873	

3.3.2 Estimation de modèles (G)ARCH : Comparaison des résultats de plusieurs logiciels (ECTS, RATS et EVIEWS)

L'étude menée par Fiorentini, Calzolari et Panattoni (1996) révèle qu'il est essentiel d'utiliser, autant que possible, les dérivées analytiques, plutôt que numérique, lors de l'estimation des modèles (G)ARCH.

Ici, nous allons comparer des résultats obtenus en utilisant les dérivées analytiques (ECTS) et les numériques (RATS et EVIEWS). Ces estimations ont abouti aux résultats suivants :

Tableau 3.9

Résultats comparatifs.

	GE	IBM	S&P500
$\hat{\alpha}_0$	1.5841(1)	3.0551(1)	0.7636(1)
	1.5808(2)	3.0370(2)	0.6369(2)
	1.5161(3)	3.1184(3)	0.6431(3)
$\hat{\alpha}_1$	0.0972(1)	0.1164(1)	0.1119(1)
	0.0949(2)	0.0965(2)	0.1152(2)
	0.0944(3)	0.0986(3)	0.1180(3)
$\hat{\delta}_1$	0.8834(1)	0.8345(1)	0.8705(1)
	0.8782(2)	0.8354(2)	0.8673(2)
	0.8798(3)	0.8318(3)	0.8649(3)
$\hat{\mu}$	1.2334(1)	1.3042(1)	0.6859(1)
	1.2394(2)	1.3060(2)	0.6848(2)
	1.2459(3)	1.3199(3)	0.6965(3)
Log	-2969.380(1)	-2905.156(1)	-2641.256(1)
vraisemblance	-2971.749(2)	-2908.199(2)	-2641.432(2)
	-2962.60(3) ²	-2897.306(3)	-2634.283(3)

Où (1), (2) et (3) représentent les résultats obtenus, respectivement avec ECTS, EVIEWS et RATS. On constate qu'il y a une forte similitude au niveau des résultats donnés par l'ensemble des trois logiciels.

² L'estimation en RATS donne la fonction value. Pour obtenir la log-vraisemblance, il faut rajouter un terme $-(1/2)\log(2\pi)$ à chaque contribution. Pour l'ensemble de l'échantillon, on a alors $-(n/2)\log(2\pi)$.

Notons que ces résultats ont été obtenus en utilisant l'algorithme BHHH, comme algorithme d'optimisation, avec un critère de convergence de 10^{-5} .

L'estimation des modèles (G)ARCH à l'aide de logiciels et de programmes différents (voir annexe), nous permet ainsi de tenir compte des recommandations des études de McCulloch et Renfro (1999), puis Brooks, Burke et Persaud (2001).

Car, selon ces auteurs l'utilisation des logiciels et programmes distincts permet d'obtenir des estimations plus fiables, pour ce genre de modèle.

Nous validons ainsi les estimations obtenues à l'aide des dérivées analytiques de la log-vraisemblance des modèles (G)ARCH par des dérivées numériques de la log-vraisemblance de ces mêmes modèles.

Conclusion

Dans le présent mémoire, nous avons présenté une classe de modèle utilisée dans l'estimation de la volatilité conditionnelle des données macroéconomiques et financières. Nous avons d'abord commencé par présenter la formulation et les propriétés de ces modèles. À cet effet, nous avons présenté en détail les modèles (G)ARCH. Et de manière brève, nous avons ensuite présenté les extensions de ces modèles. Puis, nous avons décrit en détail les propriétés de la loi normale, de laquelle on déduit les propriétés de ces modèles. Enfin, nous avons passé en revue la méthode d'estimation la plus utilisée dans le contexte de la modélisation ARCH, de même que les difficultés auxquelles on fait face lors des estimations, ainsi que les procédures de constructions de tests.

Par la suite nous avons appliqué cette méthodologie à des rendements boursiers mensuels. Nous avons remarqué qu'en plus de présenter des effets ARCH, ces données sont caractérisées par un accroissement des queues de distributions à travers le temps, phénomènes que les experts qualifient de leptokurtique.

Les difficultés posées par l'estimation, selon des travaux antérieurs, de ces modèles nous a amené à considérer une estimation par MV ayant recours à des dérivées analytiques du MV de ces modèles. Puis nous avons confronté ces résultats avec ceux obtenus à l'aide de dérivées numériques du MV de ces mêmes modèles, afin de valider les premiers.

Bien que ces modèles fournissent des résultats cohérents; il est à remarquer que la formulation ARCH repose sur des hypothèses qui peuvent s'écarter plus ou moins de la situation réelle.

En effet, les modèles ARCH et GARCH sont tout à fait symétriques, c'est-à-dire que les effets des chocs ne sont pas différenciés selon leurs signes, et pourtant l'asymétrie représente une hypothèse très réaliste pour des séries financières.

Annexe

Programmes ECTS

Tests du Skewness, du Kurtosis et de Jarque-Bera

```

sample 1 888
read m-geln.dat ge
gen y = ge
gen iota = 1

ols y iota
gen u2=res*res
gen u3=res*u2
gen u4=u2*u2
gen mom2=sum(u2)/nobs
gen mom3=sum(u3)/nobs
gen mom4=sum(u4)/nobs
set Sk=mom3(nobs)/mom2(nobs)^(3/2)
set Ku=mom4(nobs)/mom2(nobs)^2
set T=nobs/6*Sk^2+nobs/24*(Ku-3)^2
set Pvalue=1-chisq(T,2)
print Pvalue T Sk Ku

```

```
quit
```

```
#*****Fin du
```

```
programme*****
```

Résultats

Sample set from observation 1 to observation 888

```

read m-geln.dat ge
Variables ge now read in

```

```

gen y = ge
gen iota = 1
ols y iota

```

Ordinary Least Squares:

Variable Parameter estimate Standard error T statistic

iota	1.037578	0.259896	3.992280
------	----------	----------	----------

Number of observations = 888 Number of regressors = 1
 Mean of dependent variable = 1.037578
 Sum of squared residuals = 53202.970757
 Explained sum of squares = 955.991930
 Estimate of residual variance
 (with d.f. correction) = 59.980801
 Standard error of regression = 7.744727
 R squared (uncentred) = 0.017652

Estimated covariance matrix:

0.067546

```

gen u2=res*res
gen u3=res*u2
gen u4=u2*u2
gen mom2=sum(u2)/nobs
gen mom3=sum(u3)/nobs
gen mom4=sum(u4)/nobs
set Sk=mom3(nobs)/mom2(nobs)^(3/2)
set Ku=mom4(nobs)/mom2(nobs)^2
set T=nobs/6*Sk^2+nobs/24*(Ku-3)^2
set Pvalue=1-chisq(T,2)
print Pvalue T Sk Ku
Pvalue = 0.000000
T = 1078.220509
Sk = -0.252085
Ku = 8.374654
  
```

Quit

Estimation du modèle ARCH(1)

```

sample 1 888
read m-geln.dat ge
gen y = ge
gen iota=1

ols y iota
set ltilde=(nobs/2)*(log(nobs/(2*PI))-1-log(ssr))
set bt=coef(1)
gen u2=res*res
gen u2l1=lag(1,u2)
sample 2 360
ols u2 c u2l1
set a=coef(1)
set g=coef(2)
  
```

```
set lb=-20
```

```
procedure arch 3
sample 1 888
gen u=y-bt
gen ulag=lag(1,u)
gen u2=u^2
gen u2lag=lag(1,u2)
gen s2=a+g*u2lag
gen lhd=max(lb,-(log(2*PI*s2)+u2/s2)/2)
answer lhd
gen dlds2=(lhd > lb)*(u2-s2)/(2*s2^2)
gen dldu=-u/s2
gen dlda=dlds2
answer dlda
gen dldg=dlds2*u2lag
answer dldg
gen dldbt=-2*dlds2*g*ulag-dldu
answer dldbt
end
```

```
set showprogress=1
sample 3 888
```

```
mlopg arch(a,g,bt,1)
deriv a=arch(a,g,bt,2)
deriv g=arch(a,g,bt,3)
deriv bt=arch(a,g,bt,4)
end
```

```
# Vérification des conditions de premier ordre (CPO)
ols iota CG
```

```
quit
#*****Fin du
programme*****
```

Résultats

```
sample 1 888
Sample set from observation 1 to observation 888
```

```
read m-geln.dat ge
Variables ge now read in
```

```
gen y = ge
gen iota=1
ols y iota
```

Ordinary Least Squares:

Variable	Parameter estimate	Standard error	T statistic
----------	--------------------	----------------	-------------

iota	1.037578	0.259896	3.992280
------	----------	----------	----------

Number of observations = 888 Number of regressors = 1

Mean of dependent variable = 1.037578

Sum of squared residuals = 53202.970757

Explained sum of squares = 955.991930

Estimate of residual variance

(with d.f. correction) = 59.980801

Standard error of regression = 7.744727

R squared (uncentred) = 0.017652

Estimated covariance matrix:

0.067546

```
set ltilde=(nobs/2)*(log(nobs/(2*PI))-1-log(ssr))
```

```
set bt=coef(1)
```

```
gen u2=res*res
```

```
gen u2l1=lag(1,u2)
```

```
sample 2 888
```

```
Sample set from observation 2 to observation 888
```

```
ols u2 c u2l1
```

Ordinary Least Squares:

Variable	Parameter estimate	Standard error	T statistic
----------	--------------------	----------------	-------------

constant	47.939613	5.709211	8.396889
----------	-----------	----------	----------

u2l1	0.201601	0.032956	6.117242
------	----------	----------	----------

Number of observations = 887 Number of regressors = 2

Mean of dependent variable = 59.969498

Sum of squared residuals = 2.255115e+07

Explained sum of squares = 4.143489e+06

Estimate of residual variance

(with d.f. correction) = 25481.524620

Standard error of regression = 159.629335

R squared (uncentred) = 0.155218 (centred) = 0.040568

Estimated covariance matrix:

32.595095 -0.064810

-0.064810 0.001086

```

set a=coef(1)
set g=coef(2)
set lb=-20
procedure arch 3
sample 1 888
gen u=y-bt
gen ulag=lag(1,u)
gen u2=u^2
gen u2lag=lag(1,u2)
gen s2=a+g*u2lag
gen lhd=max(lb,-(log(2*PI*s2)+u2/s2)/2)
answer lhd
gen dlds2=(lhd > lb)*(u2-s2)/(2*s2^2)
gen dldu=-u/s2
gen dlda=dlds2
answer dlda
gen dldg=dlds2*u2lag
answer dldg
gen dldbt=-2*dlds2*g*ulag-dldu
answer dldbt
end
set showprogress=1
sample 3 888
Sample set from observation 3 to observation 888

```

```

mlogp arch(a,g,bt,1)
deriv a=arch(a,g,bt,2)
deriv g=arch(a,g,bt,3)
deriv bt=arch(a,g,bt,4)
end

```

Maximising a Sum of Contributions:

Number of iterations = 21

Parameter	Parameter estimate	Standard error	T statistic
-----------	--------------------	----------------	-------------

a	43.105132	1.744441	24.710002
g	0.216124	0.031784	6.799700
bt	1.209378	0.228865	5.284241

Number of observations = 886 Number of estimated parameters = 3

Maximised value of criterion function = -3043.694986

Estimated covariance matrix:

3.043073	-0.027216	0.048408
-0.027216	0.001010	-0.000401
0.048408	-0.000401	0.052379

ols iota CG

Ordinary Least Squares:

Variable	Parameter estimate	Standard error	T statistic
CG1	-0.000034	1.747402	-0.000019
CG2	-1.634001e-08	0.031838	-5.132189e-07
CG3	5.600483e-07	0.229253	0.000002

Number of observations = 886 Number of regressors = 3

Mean of dependent variable = 1.000000

Sum of squared residuals = 886.000000

Explained sum of squares = 0.000000

Estimate of residual variance

(with d.f. correction) = 1.003398

Standard error of regression = 1.001697

R squared (uncentred) = 0.000000

Estimated covariance matrix:

3.053412	-0.027308	0.048573
-0.027308	0.001014	-0.000403
0.048573	-0.000403	0.052557

quit

Estimation du modèle GARCH(1,1)

sample 1 888

read m-geln.dat ge

gen y = ge

gen iota = 1

sample 2 888

ols y iota

set ltilde = (nobs/2)*(log(nobs/(2*PI))-1-log(ssr))

set bt = coef(1)

gen u2 = res*res

gen u2l1 = lag(1,u2)

gen u2l2 = lag(2,u2)

initialisation des valeurs de départ

sample 2 888

ols u2 c u2l1 u2l2

set a = abs(coef(1))

```

set g = abs(coef(2))
set d = abs(coef(3))

# imposons les bornes inférieure afin d'éviter des problèmes numériques.
set lb = -20
gen t = time(0)

# définition des dérivées(cpo)
procurement garch 4
gen u = y-bt
gen u2 = u^2
gen D = d^(t-1)
gen ss = lag(1,conv(D,u2))
gen s2 = a/(1-d)+a*g*D/((1-g-d)*(1-d))+g*ss
gen lhd = max(lb,-(log(2*PI*s2)+u2/s2)/2)
answer lhd
gen dlds2 = (lhd>lb)*(u2-s2)/(2*s2^2)
gen dllda = dlds2*(1/(1-d)+g*D/((1-g-d)*(1-d)))
answer dllda
gen dlldg = dlds2*(a*D/((1-g-d)^2)+ss)
answer dlldg
gen dt = (t-1)*lag(1,D)
gen dd = a*g*dt/((1-g-d)*(1-d))
gen dd = dd+a*g*D*(2-g-2*d)/((1-g-d)^2*(1-d)^2)
gen dlidd = dlds2*(a/(1-d)^2+dd+g*lag(1,conv(dt,u2)))
answer dlidd

gen dlldbt = u/s2-2*g*dlds2*lag(1,conv(D,u))
answer dlldbt
end

# fonction de vraisemblance
mlopg garch(a,g,d,bt,1)
deriv a = garch(a,g,d,bt,2)
deriv g = garch(a,g,d,bt,3)
deriv d = garch(a,g,d,bt,4)
deriv bt = garch(a,g,d,bt,5)
end

# Vérification des conditions de premier ordre (CPO)
ols iota CG

quit
#*****Fin du
programme*****

```

Résultats

sample 1 888

Sample set from observation 1 to observation 888

```
read m-geln.dat ge
Variables ge now read in
```

```
gen y = ge
gen iota = 1
sample 2 888
Sample set from observation 2 to observation 888
```

```
ols y iota
```

Ordinary Least Squares:

Variable	Parameter estimate	Standard error	T statistic
----------	--------------------	----------------	-------------

iota	1.034008	0.260165	3.974436
------	----------	----------	----------

Number of observations = 887 Number of regressors = 1

Mean of dependent variable = 1.034008

Sum of squared residuals = 53192.933223

Explained sum of squares = 948.355848

Estimate of residual variance

(with d.f. correction) = 60.037171

Standard error of regression = 7.748366

R squared (uncentred) = 0.017516

Estimated covariance matrix:

0.067686

```
set ltilde = (nobs/2)*(log(nobs/(2*PI))-1-log(ssr))
```

```
set bt = coef(1)
```

```
gen u2 = res*res
```

```
gen u2l1 = lag(1,u2)
```

```
gen u2l2 = lag(2,u2)
```

```
# initialisation des valeurs de départ
```

```
sample 2 888
```

```
Sample set from observation 2 to observation 888
```

```
ols u2 c u2l1 u2l2
```

Ordinary Least Squares:

Variable	Parameter estimate	Standard error	T statistic
----------	--------------------	----------------	-------------

constant	37.850777	5.797730	6.528551
----------	-----------	----------	----------

u2l1	0.158765	0.032904	4.825069
------	----------	----------	----------

u2l2	0.212083	0.032902	6.445821
------	----------	----------	----------

Number of observations = 887 Number of regressors = 3

Mean of dependent variable = 59.969485
 Sum of squared residuals = 2.153721e+07
 Explained sum of squares = 5.155946e+06
 Estimate of residual variance
 (with d.f. correction) = 24363.364283
 Standard error of regression = 156.087681
 R squared (uncentred) = 0.193156 (centred) = 0.083648

Estimated covariance matrix:

```

33.613671 -0.051548 -0.051503
-0.051548 0.001083 -0.000219
-0.051503 -0.000219 0.001083
  
```

```

set a = abs(coef(1))
set g = abs(coef(2))
set d = abs(coef(3))
# imposons les bornes inférieure afin d'éviter des problèmes numériques.
set lb = -20
gen t = time(0)
# définition des dérivées(cpo)
procedure garch 4
gen u = y-bt
gen u2 = u^2
gen D = d^(t-1)
gen ss = lag(1,conv(D,u2))
gen s2 = a/(1-d)+a*g*D/((1-g-d)*(1-d))+g*ss
gen lhd = max(lb,-(log(2*PI*s2)+u2/s2)/2)
answer lhd
gen dlds2 = (lhd>lb)*(u2-s2)/(2*s2^2)
gen dlda = dlds2*(1/(1-d)+g*D/((1-g-d)*(1-d)))
answer dlda
gen dldg = dlds2*(a*D/((1-g-d)^2)+ss)
answer dldg
gen dt = (t-1)*lag(1,D)
gen dd = a*g*dt/((1-g-d)*(1-d))
gen dd = dd++a*g*D*(2-g-2*d)/((1-g-d)^2*(1-d)^2)
gen dlld = dlds2*(a/(1-d)^2+dd+g*lag(1,conv(dt,u2)))
answer dlld
gen dlldbt = u/s2-2*g*dlds2*lag(1,conv(D,u))
answer dlldbt
end
# fonction de vraisemblance
mlogp garch(a,g,d,bt,1)
deriv a = garch(a,g,d,bt,2)
deriv g = garch(a,g,d,bt,3)
deriv d = garch(a,g,d,bt,4)
deriv bt = garch(a,g,d,bt,5)
end
  
```

Maximising a Sum of Contributions:

Number of iterations = 17

Parameter	Parameter estimate	Standard error	T statistic
-----------	--------------------	----------------	-------------

a	1.584184	0.539653	2.935559
g	0.097273	0.020630	4.715240
d	0.883428	0.021313	41.449811
bt	1.233484	0.202915	6.078824

Number of observations = 887 Number of estimated parameters = 4
 Maximised value of criterion function = -2969.380013

Estimated covariance matrix:

0.291226	0.005294	-0.008946	0.000591
0.005294	0.000426	-0.000398	-0.000076
-0.008946	-0.000398	0.000454	0.000085
0.000591	-0.000076	0.000085	0.041174

ols iota CG

Ordinary Least Squares:

Variable	Parameter estimate	Standard error	T statistic
----------	--------------------	----------------	-------------

CG1	-0.004646	0.540874	-0.008590
CG2	-0.000185	0.020676	-0.008940
CG3	0.000231	0.021361	0.010815
CG4	-0.000193	0.203374	-0.000949

Number of observations = 887 Number of regressors = 4

Mean of dependent variable = 1.000000

Sum of squared residuals = 886.999870

Explained sum of squares = 0.000130

Estimate of residual variance

(with d.f. correction) = 1.004530

Standard error of regression = 1.002262

R squared (uncentred) = 1.470073e-07

Estimated covariance matrix:

0.292545	0.005318	-0.008986	0.000594
0.005318	0.000428	-0.000400	-0.000076
-0.008986	-0.000400	0.000456	0.000085
0.000594	-0.000076	0.000085	0.041361

quit

Fonction de densité cumulative

set savegnu = 1

sample 1 888

read m-geln.dat ge

gen y = ge

gen iota = 1

ols y iota

set ect = sqrt(errvar)

gen u = res/ect

set linestyle = 1

gen x = -4 + 8*time(-1)/888

gen w = random()

mat edft = cdf(u,x)

mat lsnorm = cdf(w,x)

plot (x lsnorm edft)

quit

#*****Fin du

programme*****

Programmes RATS

```

* Estimation d'un ARCH(1)
all 0 888:1
open data m-ibmspgeln.xls
data(org=obs,format=xls) / ibm sp500 ge
*
*
* Régression
linreg ge / res_ge
# constant
*
*
* Test d'homoscédasticité
set uge = res_ge
set u2ge = uge**2
*
linreg u2ge /
# constant u2ge{1}
*
compute chistat=%nobs*%rsquared
cdf chisq chistat 1
*
*
declare series u
declare series h
nonlin vc vb bc
frml resid = ge(t) - bc
frml archvar = vc+vb*u{1}**2
frml archlog =(h(T)=archvar(T)),(u(T))= resid(T),-0.5*(log(h(T))+u(T)**2/h(T))
linreg ge / u
# constant
compute bc=%beta(1)
compute vc=%seesq , vb=0.05
set h = %seesq
maximize(method=simplex,itors=5) archlog
maximize(method=bhhh,robust,itors=50,vcv) archlog
set fv = archvar(t)
set resi = resid(t)/sqrt(fv(t))
set resisq = resi(t)*resi(t)
correlate(qstats,number=20,span=10) resi
correlate(qstats,number=20,span=10) resisq

*****Fin du
programme*****

```

Linear Regression - Estimation by Least Squares
Dependent Variable GE

Usable Observations 888 Degrees of Freedom 887
 Centered R**2 -0.000000 R Bar **2 -0.000000
 Uncentered R**2 0.017652 T x R**2 15.675
 Mean of Dependent Variable 1.0375777027
 Std Error of Dependent Variable 7.7447273228
 Standard Error of Estimate 7.7447273228
 Sum of Squared Residuals 53202.970757
 Durbin-Watson Statistic 1.860888

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	1.0375777027	0.2598960319	3.99228	0.00007085

Linear Regression - Estimation by Least Squares

Dependent Variable U2GE

Usable Observations 887 Degrees of Freedom 885
 Centered R**2 0.040568 R Bar **2 0.039484
 Uncentered R**2 0.155218 T x R**2 137.678
 Mean of Dependent Variable 59.96949778
 Std Error of Dependent Variable 162.87722121
 Standard Error of Estimate 159.62933509
 Sum of Squared Residuals 22551149.289
 Regression F(1,885) 37.4206
 Significance Level of F 0.00000000
 Durbin-Watson Statistic 2.083255

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	47.939613325	5.709211431	8.39689	0.00000000
2. U2GE{1}	0.201601201	0.032956226	6.11724	0.00000000

Chi-Squared(1)= 35.983706 with Significance Level 0.00000000

Linear Regression - Estimation by Least Squares

Dependent Variable GE

Usable Observations 888 Degrees of Freedom 887
 Centered R**2 -0.000000 R Bar **2 -0.000000
 Uncentered R**2 0.017652 T x R**2 15.675
 Mean of Dependent Variable 1.0375777027
 Std Error of Dependent Variable 7.7447273228
 Standard Error of Estimate 7.7447273228
 Sum of Squared Residuals 53202.970757
 Durbin-Watson Statistic 1.860888

Variable	Coeff	Std Error	T-Stat	Signif

1. Constant	1.0375777027	0.2598960319	3.99228	0.00007085

MAXIMIZE - Estimation by Simplex

Usable Observations 887
 Total Observations 888 Skipped/Missing 1
 Function Value -2234.50751863

Variable	Coeff

1. VC	48.642029550
2. VB	0.176000167
3. BC	1.160528599

MAXIMIZE - Estimation by BHHH

Convergence in 12 Iterations. Final criterion was 0.0000064 < 0.0000100
 Usable Observations 887
 Total Observations 888 Skipped/Missing 1
 Function Value -2234.49236258

Variable	Coeff	Std Error	T-Stat	Signif

1. VC	48.587529486	1.531533416	31.72476	0.00000000
2. VB	0.182981783	0.030043945	6.09047	0.00000000
3. BC	1.151777081	0.259536004	4.43783	0.00000909

Covariance\Correlation Matrix of Coefficients

	VC	VB	BC
VC	2.34559460535	-0.4398104729	0.1524096960
VB	-0.02023713392	0.00090263865	-0.0700483941
BC	0.06058103482	-0.00054620134	0.06735893735

Correlations of Series RESI

Autocorrelations

1: 0.0365912 -0.0247701 -0.0780971 0.0507669 0.0317496 -0.0313592
 7: 0.0243762 0.0459111 0.0239507 -0.0141114 0.0222968 0.0756906
 13: 0.0103144 -0.0509369 -0.0102348 -0.0232884 0.0710716 0.0398931
 19: -0.0307200 0.0002483

Ljung-Box Q-Statistics

Q(10) = 14.3795. Significance Level 0.15637570

Q(20) = 29.8950. Significance Level 0.07157343

Correlations of Series RESISQ

Autocorrelations

1: -0.0071496 0.1201823 0.1521879 0.1226490 0.0631870 0.1035382
 7: 0.1534080 0.0961647 0.2228081 0.1505662 0.1551432 0.1358145
 13: 0.0629790 0.0603818 0.1480355 0.0434628 0.0891030 0.0303276
 19: 0.2660957 0.0401298

Ljung-Box Q-Statistics

Q(10) = 154.5266. Significance Level 0.00000000

Q(20) = 295.0356. Significance Level 0.00000000

* Estimation du GARCH(1,1)

all 0 888:1

open data m-ibmspgeln.xls

data(org=obs,format=xls) / ibm sp500 ge

*

*

* Régression

linreg ge / res_ge

constant

*

*

* Test d'homoscédasticité

set uge = res_ge

set u2ge = uge**2

*

linreg u2ge /

constant u2ge{1}

*

compute chistat=%nobs*%rsquared

cdf chisq chistat 1

*

*

* Estimation d'un GARCH(1,1)

declare series u

declare series h

nonlin vc va vb bc

frml resid = ge - bc

frml hf = vc+va*h{1}+vb*u{1}**2

frml loggarch =(h(T)=hf(T)),(u(T))= resid(T),-0.5*(log(h(T))+u(T)**2/h(T))

linreg ge / u

constant

compute bc=%beta(1)

compute vc=%seesq , va=0.4 , vb=0.3

```
set h = %seesq
maximize(method=simplex, iters=5) loggarch 5 888
maximize(method=bhhh, robust, iters=50, vcv) loggarch 5 888
```

```
*****Fin du
programme*****
```

Linear Regression - Estimation by Least Squares

Dependent Variable GE

```
Usable Observations 888 Degrees of Freedom 887
Centered R**2 -0.000000 R Bar **2 -0.000000
Uncentered R**2 0.017652 T x R**2 15.675
Mean of Dependent Variable 1.0375777027
Std Error of Dependent Variable 7.7447273228
Standard Error of Estimate 7.7447273228
Sum of Squared Residuals 53202.970757
Durbin-Watson Statistic 1.860888
```

Variable	Coeff	Std Error	T-Stat	Signif

1. Constant	1.0375777027	0.2598960319	3.99228	0.00007085

Linear Regression - Estimation by Least Squares

Dependent Variable U2GE

```
Usable Observations 887 Degrees of Freedom 885
Centered R**2 0.040568 R Bar **2 0.039484
Uncentered R**2 0.155218 T x R**2 137.678
Mean of Dependent Variable 59.96949778
Std Error of Dependent Variable 162.87722121
Standard Error of Estimate 159.62933509
Sum of Squared Residuals 22551149.289
Regression F(1,885) 37.4206
Significance Level of F 0.00000000
Durbin-Watson Statistic 2.083255
```

Variable	Coeff	Std Error	T-Stat	Signif

1. Constant	47.939613325	5.709211431	8.39689	0.00000000
2. U2GE{1}	0.201601201	0.032956226	6.11724	0.00000000

Chi-Squared(1)= 35.983706 with Significance Level 0.00000000

Linear Regression - Estimation by Least Squares

Dependent Variable GE

```
Usable Observations 888 Degrees of Freedom 887
Centered R**2 -0.000000 R Bar **2 -0.000000
Uncentered R**2 0.017652 T x R**2 15.675
Mean of Dependent Variable 1.0375777027
```

Std Error of Dependent Variable 7.7447273228
 Standard Error of Estimate 7.7447273228
 Sum of Squared Residuals 53202.970757
 Durbin-Watson Statistic 1.860888

Variable	Coeff	Std Error	T-Stat	Signif

1. Constant	1.0375777027	0.2598960319	3.99228	0.00007085

MAXIMIZE - Estimation by Simplex
 Usable Observations 884
 Function Value -2149.30384176

Variable	Coeff

1. VC	1.8407540960
2. VA	0.8513577346
3. VB	0.1112519495
4. BC	0.9430103182

MAXIMIZE - Estimation by BHHH
 Convergence in 15 Iterations. Final criterion was 0.0000000 < 0.0000100
 Usable Observations 884
 Function Value -2146.58435830

Variable	Coeff	Std Error	T-Stat	Signif

1. VC	1.5161693395	0.5546314801	2.73365	0.00626362
2. VA	0.8798204045	0.0224236475	39.23628	0.00000000
3. VB	0.0944697165	0.0189528601	4.98446	0.00000062
4. BC	1.2459900757	0.2035170009	6.12229	0.00000000

Covariance\Correlation Matrix of Coefficients

	VC	VA	VB	BC
VC	0.30761607874	-0.7581076753	0.4033610532	-0.0016264207
VA	-0.00942847964	0.00050281997	-0.8797642370	0.0202348661
VB	0.00424007203	-0.00037389299	0.00035921090	-0.0106305269
BC	-0.00018358538	0.00009234370	-0.00004100438	0.04141916966

Documents de travail et articles trouvés sur Internet

Journal of Applied Econometrics :

<http://jae.wiley.com/jae/>

Journal of Econometrics:

<http://www.elsevier.com:80/cgi-bin/cas/tree/store/econom/casfree/browse/browse.cgi>

Site ECTS:

<http://russell.cnrs-mrs.fr/pub/ects3/aide/liens.htm>

Site RATS :

<http://www.estima.com>

Bibliographie

BAILLIE, R. T, and DeGENNARO, R.P. « Stock Return and volatility », *Journal of Financial and Quantitative Analysis*, 25(2), 1990, p.203-214.

BERA A.K. & M. L. HIGGINS. « ARCH models : Propreties, Estimations and Testing », *Journal of Econometric Survey*, 7(4), 1993, p.305-366.

BOLLERSLEV T., R. CHOU & K. KRONER. « ARCH modelling in finance: A Review of theory and empirical evidence », *Journal of Econometrics*, 52, 1992, p.5-59.

BOLLERSLEV, ENGLE et NELSON. « ARCH models », *Handbook of Econometrics*, vol. IV, Elsevier, 1994.

BOLLERSLEV, T. « A Conditional Heteroscedastic Time Series Model for Speculative and rates of Return », *Review of Economics and Statistics*, 69, p.542-547.

BOLLERSLEV, T. « A Conditional Heteroskedasticity », *Journal of Econometrics*, 31, 1986, p.309-328.

BOLLERSLEV, T. et WOOLDRIDGE, J.M. « *Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances* », *Econometric review*, 11(2), 1992, p.143-172.

BOUGEROL et PICARD. « Stationarity of GARCH processus and some Nonnegative Time Series », *Journal of Econometrics*, 52, 1990, p.115-127.

BROOKS, Chris. *Introductory econometrics for finance* , Cambridge University Press, 2002.

CHOU, R.Y. « Volatility Persistence and Stocks Valuations », *Journal of Applied Econometrics*, 3, 1988, p.279-294.

DAVIDSON, R. et MAC KINNON, J.G. *Econometric Theory and Methods*, Oxford University Press New York, 2004.

DAVIDSON, R. et MAC KINNON, J.G. *Estimation and inference in econometrics*, Oxford University Press New York, 1993.

DROESBEKE, J.J. FICHET, B. TASSI, P. *Modélisation ARCH, théorie statistique et application dans le domaine de la finance*, Éditions de l'université de Bruxelles, 1994, 242 p.

ENGLE, R.F and BOLLERSLEV, T. « Modelling the persistence of conditional variances », *Econometric Reviews*, 5, 1986, p.1-50.

ENGLE, R.F. « Autoregressive Conditional Heteroskedasticity with estimates of the variance of United Kingdom inflation », *Econometrica*, 50(4), 1982, p.987-1007.

FIORENTINI, G. , CALZOLARI, G. et PANATTONI, L. « Analytic derivatives and computation of GARCH estimates », *Journal of applied econometrics*, vol.11, 1996, p.399-417.

FRENCH, K.R. SCHWERT, G.W. et STAMBAUCH, R.F. « Expected stock returns and volatility », *Journal of financial Economics*, 19, 1987, p.3-30.

GLOSTEN, JAGANNATHON and RUNKLE. « On the relation between the expected value and the volatility of the nominal excess returns on stocks », *Journal of finance*, 1993.

GOURIEROUX Christian et Alain MONFORT. *Modèles ARCH et Applications Financières*, Economica, 1992.

HAMILTON, James D. « A new approach to the economic analysis of nonstationarity time series and the business cycle », *Econometrica*, 57, 1989, p.357-384.

HAMILTON, James D. *Times series analysis*, Princeton University Press, 1994.

JARQUE, C.M et BERA, A.K. « Efficient test for normality, heteroscedasticity and a serial independence of regression residuals », *Economic Letters*, vol.6, 1980, p.29-61.

KENDALL, M and STUART, A. *The advanced theory of statistics*, vol.1, 4^{ième} edition, MacMillan Publishing Co.,Inc, 1977.

KWIATKOWSKI, PHILLIPS, SCHMIDT and SHIN. « Testing the null hypothesis of stationarity against the alternative of unit root : how sure are we that economic series have a unit root ? », *Journal of Econometrics*, 54, 1992, p.159-178.

LARDIC, S et MIGNON, V. *Économétrie des Séries Temporelles Macroéconomiques et Financières*, Economica, 2002.

MANDELBROT, B. « The variation of Certain Prices », *J.Business*, 1963, p.394-419.

NELSON, Daniel. « Conditional Heteroskedasticity in asset returns : A new approach », *Econometrica*, 59, 1991, p.347-370.

NELSON, Daniel. « Stationarity and persistence in GARCH(1,1) model », *Econometric Theory*, vol.6, 1990b, p.318-338.

PATEL, J.K and READ, C.B. *Handbook of the normal distribution*, vol.40, Marcel Dekker Inc, 1982.

SCHWARZ, G William and SEGIN, Paul J. « Heteroskedasticity in stock returns », *Journal of finance*, 45, 1990, p.1129-1155.

TSAY, R.S. « Conditional heteroscedastic time series models », *Journal of the American Statistical Association*, vol.82, No 398, Juin 1987, p.590-604.